# Determinism and Responsibility

## A Neuropsychological Defence of Criminal Liability

Alexander James Collins Hinchliffe, LLB (Hons), LLM, MSc, Barrister

Tesis Doctoral UPF / 2022

DIRECTORES DE TESIS

Prof. Dr. Josep Joan Moreso / Prof. Dr. Rosemarie Nagel / Dr. Ivó Coca Vila

DEPARTAMENTO DE DERECHO

Universitat Pompeu Fabra Barcelona

# Acknowledgments

I would like to thank each of my thesis supervisors – Josep Joan Moreso, for your unwavering faith, support and guidance throughout this six-year project; and Rosemarie Nagel and Ivó Coca Vila, for taking a leap of faith in joining the project midway and offering your invaluable oversight and advice.

I would also like to thank my husband, Danny Hinchliffe, who has been very, *very* patient.

# Abstract

The philosophical debate concerning free will and determinism has continued unresolved for millennia, the outcome of which allegedly carries critical implications for whether and how people are held responsible for their actions. The law generally presumes the existence of some manner of free will and constructs the concept of responsibility on top of these foundations. Without entering directly into the philosophical debate, the present thesis adopts the opposing presumption that metaphysical free will is precluded by causal determinism and therefore does not exist. From this starting position, the question asked is *how can people rationally be held responsible for their actions in a deterministic universe absent of metaphysically free decision-making?* Part One of the thesis reviews an extensive body of empirical and theoretical research concerning decision-making in the human brain from the joint studies of neuroscience and psychology. The reasoning for this is twofold; on the one hand, and in very general terms, neuropsychology is broadly premised on the same presumptions of causal determinism which underpin the natural sciences – *i.e.,* the physics of the macroscopic universe, chemistry, biology, and the biochemistry governing biological organisms. On the other hand, every action which results in criminal liability begins with a decision to act in the brain, whether occurring consciously or unconsciously.

Part Two of the thesis subsequently takes conclusions and implications from the neuropsychology of decision-making and applies them to a critique of the legal concept of responsibility, focusing on the aspect of *mens rea* in particular. The thesis proposes replacing proof of subjective states of mind with proof of certain mental capacities that are necessary for responsibility – the capacities to exercise ordinary self-control, to recognise and respond to reason, and to understand the nature and consequences of one's actions. This approach is tested against leading jurisprudence to demonstrate its efficacy as a practical means of ascribing legal responsibility for criminal actions. Further implications are presented, such as replacing the notion of moral blame with that of unreasonable conduct as the theoretical underpinning to criminal liability, and denying any role for retributive theories of punishment. Finally, this capacity-based theory of responsibility is reintroduced into the wider philosophical debate between free will and determinism. Here, it is demonstrated how the theory provides a rational approach to

holding people responsible for their actions in a deterministic universe and without any presumptions of, or reliance upon, metaphysical free will.

*

## Resumen

El debate filosófico sobre el libre albedrío y el determinismo, pese a su larga tradición, dista mucho de estar resuelto. Este, sin embargo, tendría importantes consecuencias a la hora de determinar si y cómo las personas habrían de ser responsabilizadas por sus acciones. El Derecho presume con carácter general la existencia de una suerte de libre albedrío y construye sobre tal premisa su concepto legal de responsabilidad. Aunque sin ocuparme directamente del debate filosófico, la presente tesis adopta como punto de partida la presunción metafísica opuesta al libre albedrío, esto es, la determinista. A partir de aquí, la cuestión central es la siguiente: *¿cómo puede ser hecha responsable de forma racional una persona por sus acciones en un universo determinado en el que no existe una libertad metafísica para tomar decisiones?* En la primera parte de la tesis se analiza la literatura empírica y teórica sobre los procesos de toma de decisiones en la mente humana con base en investigaciones neurocientíficas y psicológicas. De ahí se extraen las dos siguientes conclusiones. Por un lado, y en términos muy genéricos, la neuropsicología está ampliamente sustentada sobre las mismas presunciones del determinismo causal que subyacen a las ciencias naturales, por ejemplo, a la física del universo microscópico, la química, la biología o la bioquímica que regula los organismos biológicos. Por otro lado, toda acción penalmente responsable comienza con una decisión de actuación en el cerebro, ya sea consciente o inconsciente.

En la segunda parte de la tesis se extraen las necesarias conclusiones e implicaciones de la neuropsicología de la toma de decisiones para una crítica al concepto legal de responsabilidad, en especial, al concepto tradicional de mens rea. En esta tesis se propone reemplazar la prueba de los estados mentales subjetivos por la prueba de unas ciertas capacidades mentales que son condición necesaria para la responsabilidad – las capacidades para ejercer un autocontrol ordinario, para reconocer y responder a una razón, así como para entender la naturaleza y las consecuencias de una acción. Este

planteamiento es sometido a examen de la mano de la jurisprudencia más relevante a fin de demostrar su eficacia y sus consecuencias prácticas a la hora de adscribir responsabilidad penal. Asimismo, se presentan ulteriores implicaciones de este planteamiento, por ejemplo, la sustitución de la noción de reproche moral por la de conducta irrazonable, una nueva aproximación a los fundamentos teóricos de la responsabilidad penal, o el rechazo de cualquier teoría retributiva del castigo. Finalmente, la teoría de la responsabilidad basada en la noción de capacidad aquí defendida es reintroducida en el amplio debate filosófico sobre el libre albedrío y el determinismo. Con ello se logra demostrar cómo dicha teoría ofrece una aproximación racional a la responsabilización de las personas por sus acciones en un mundo determinado al margen de toda presunción metafísica librealbitrista.

# 1. Introduction

> '[S]cience itself will teach man (though to my mind it's a superfluous luxury) that he never has really had any caprice or will of his own, and that he himself is something of the nature of a piano-key or an organ-stop and that there are, besides, things called the Laws of Nature; so that everything he does is not done by his willing it, but is done of itself, by the Laws of Nature. Consequently, we have only to discover these Laws of Nature, and man will no longer have to answer for his actions and life will become exceedingly easy for him. All human actions will then, of course, be tabulated according to these laws, mathematically, like tables of logarithms up to 108,000, and entered in an index; or, better still, there would be published certain edifying works of the nature of encyclopedic dictionaries, in which everything will be so clearly calculated and explained that there will be no more incidents or adventures in the world.'
>
> - Fyodor Dostoevsky, 1864.[1]

For which of our decisions and actions ought we be held responsible in a deterministic universe? For some, it is believed that a common belief in free will is necessary for the general moral fortitude of society; that a widespread belief in the absence of free will results in moral decay, or increased 'immoral' behaviour such as cheating.[2] From a legal perspective, the absence of free will has been tentatively questioned in jurisprudence. The UK House of Lords has expressed how 'the criminal law generally assumes the existence of free will. The law recognises certain exceptions, in the case of the young, those who for any reason are not fully responsible for their actions, and the vulnerable, and it acknowledges situations of duress and necessity, as also of deception and

[1] Fyodor Dostoevsky, *Notes from the Underground* (Garnett C. (trs.), Guignon C. and Aho K. (eds.), Hackett Publishing Company 2009), 18 – 19.
[2] Kathleen D. Vohs and Jonathan W. Schooler, 'The value of believing in free will: encouraging a belief in determinism increases cheating' (2008) 19(1) *Psychological Science* 49.

mistake.'[3] In the US, the Supreme Court has similarly considered that 'it is as universal and persistent in mature systems of law as belief in freedom of the human will and a consequent ability and duty of the normal individual to choose between good and evil.'[4]

Burns and Swerdlow[5] describe a 40-year-old male patient without history of social or marital problems who suddenly developed a fixation with pornography emphasising children and adolescents, solicited prostitution, and made sexual advances towards his step-daughter. Although attempting to conceal his behaviour, which he recognised as being morally and socially unacceptable, he continued to act on these new impulses stating that 'the pleasure principle overrode' his urge for restraint; he was therefore 'unable to inhibit sexual urges despite preserved moral knowledge.'[6] Following his conviction for child molestation and remittance to a 12-step rehabilitation program, he continued to be unable to restrain from soliciting staff and other patients for sexual favours, and expressed fears that he would commit rape and suicide. The patient was subsequently diagnosed with a tumour in the right orbitofrontal region of the brain – an area associated with the regulation of social behaviour and moral judgment. Following exorcise of the tumour the patient completed his treatment program and was able to return home within seven months. One year later, however, he had started collecting pornography again and complained of headaches. A brain scan revealed part of the tumour which had regrown, and his symptoms duly disappeared following further surgery.[7]

There is no question that the existence of the patient's tumour was neither his fault nor responsibility. It might equally be contended that the patient's subsequent behaviour was not his fault or responsibility either; on each occurrence, the tumour was shown not only

[3] *R v Kennedy* [2007] UKHL 38, [14].
[4] *Morissette v United States* (1952) 342 US 246, 341.
[5] Jeffrey M. Burns and Russell H. Swerdlow, 'Right orbitofrontal tumor with pedophilia symptom and constructional apraxia sign' (2003) 60(3) *Archives of Neurology* 437.
[6] *Ibid*., 437.
[7] See also David Eagleman, 'The brain on trial' (*The Atlantic*, Jul/Aug 2011) <https://www.theatlantic.com/magazine/archive/2011/07/the-brain-on-trial/308520/> accessed 15 January 2022; see also the case of Charles Joseph Whitman – the "Texas Tower Sniper" – whose actions were potentially partially linked to a tumour pressing on his amygdala, an area of the brain associated with the regulation of emotional responses including anxiety, fear and aggression.

to cause severe changes to the patient's personality, but further affected his ability to regulate behaviour and inhibit his actions, despite remaining aware that his behaviour was morally, socially and legally wrong. His personality changes *and* actions were demonstrably shown to be caused by factors entirely outside of his control. What if this case could be taken as an analogy for human decision-making and behaviour more generally; suppose that all decisions were the result of causes outside of individual control. More broadly, suppose that every decision and all behaviour arise as a combined result of each individual's genetics, epigenetics, and experiences, each being factors that are largely, if not entirely, outside of each individual's control. Suppose there is no such thing as free will? For the purposes of the present thesis, *metaphysical* free will refers to either or both of the philosophical claims that: a) it is possible for a brain to make a different decision to that which it would otherwise make when faced with the same decision in identical conditions (*i.e.,* the "principle of alternative possibilities"), or that; b) it is possible for a brain to make a decision that is completely independent of any prior causes (*i.e.,* a decision that is an "original," uncaused cause, or *causa sui*).[8]

Indeed, some of these suppositions are not particularly contentious or peculiar from the perspective of cognitive and neurosciences. Although the absence of free will is by no means universally held within these fields, it is a generally accepted proposition that there are no uncaused causes; that the macroscopic universe and all objects within it operate through the process of cause and effect; and, more specifically, that two identical causes arising within identical conditions will produce the same resultant effects. Consequently, the decisions and behaviour that the brain produces are, ultimately, caused by prior conditions and events, and there is no homunculus in the brain which operates outside of the principles of causation. It might be stated more broadly that, whilst the physics of quantum mechanics (which governs the behaviour of atoms and subatomic particles) is inherently probabilistic, the remaining sciences and the macroscopic universe that they govern are deterministic. In particular, the physics of the macroscopic universe, chemistry, biology and biochemistry are fundamentally deterministic and so, it stands as a matter of

[8] Robert Kane, 'The contours of contemporary free-will debates (Part 2)' in Kane R. (ed.), *The Oxford Handbook of Free Will* (2nd ed. Oxford University Press 2011), 5.

logic, that human psychology and behaviour is too, ultimately, constrained by the physical laws of cause and effect.

Briefly for the purposes of this introduction, an individual is generally held to be criminally responsible when they have committed a prohibited act or omission (*actus reus*)[9] with the requisite subjective state of mind (*mens rea*) such as intention, recklessness or knowledge *etc.*,[10] and in the absence of any exculpatory or justificatory factors such as coercion and duress, lack of consciousness (automatism) and self-defence *etc*.[11] Additionally, all adults are afforded the rebuttable presumption of volition – that is, the law presumes (unless proven otherwise) that all adults possess the capacity to consciously control their decisions and resulting actions and, in so doing, possess the further capacity to recognise and respond to good reasons for acting or not acting, such as the fact that a given action is legally prohibited. Whereas volition generally need not be explicitly proven by the prosecution, the negation of these capacities is the ultimate conclusion of many of the available legal defences which may be raised by defendants, such as automatism and diminished responsibility. Proof of *mens rea* is critical, however, because this reflects the underlying assumptions of the law regarding free will. That is, subjective mental states are deemed to be relevant because they reflect the dual facts that an individual has (freely) chosen to pursue a particular criminal course of conduct and, (being volitional), that they could have chosen to do otherwise.

The present thesis departs from the orthodox legal presumption of free will and aligns instead with the deterministic scientific perspective – that is to say, the thesis adopts the presumption that metaphysical free will *does not* exist, and that all human decisions and behaviour ultimately result from deterministic (and, potentially, indeterministic quantum) processes.[12] From here, the thesis investigates how and why some people may be held responsible for their decisions and actions and others are not, whilst all such decisions and actions arise from deterministic processes such that none are "freely" chosen in the

[9] Jeremy Horder, *Ashworth's Principles of Criminal Law* (9th ed. Oxford University Press 2019), Ch. 5.
[10] *Ibid*., Ch. 6.
[11] *Ibid*., Chs. 5 – 7.
[12] In this latter regard, section 13.2.4 of the thesis sets out why quantum indeterminism in the brain offers no greater route to metaphysical free will than determinism itself.

metaphysical sense. The thesis explores the processes involved in human decision-making through the lens of neuroscience and psychology; experiments and surrounding discussion from these disciplines are used to develop a scientific description of how the human brain arrives at (criminal) decisions, highlighting the implications of this description for the deterministic nature of human decision-making and actions. Next, taking as a model the method of ascribing criminal responsibility within the English common law system, the thesis investigates whether and to what extent each component of that model is rational and efficacious when considered against the scientific (and deterministic) description of human decision-making and the implications flowing therefrom.

The first part of the thesis begins with a discussion and expansion of current theories of decision-making, specifically disambiguating a single decision into five constituent parts – *what* to do, *how* to do it, *when* to do it, *whether* or not to do it, and *why* to do it. The thesis explores each of these components in turn, providing an overview of the scientific state of the art in each area and presenting key lines of experimental research which contribute to the current understanding of human decision-making. The philosophical implications of these experiments and subsequent decision-making models are discussed, and relevant links are drawn from these implications to the method of ascribing criminal responsibility within the English common law system. A central argument of the thesis is that the legal concept of *mens rea* provides an ultimately unsafe means of establishing who ought and ought not to be held responsible for their actions, and that it is rationally inconsistent with a deterministic view of human decision-making and behaviour.

The thesis proceeds to reformulate *mens rea*, replacing generally the concept of moral wrongdoing with "(un)reasonableness", and specifically introducing hybrid objective / subjective interpretations of each form of *mens rea* (*i.e.,* intention, recklessness, knowledge *etc.*). The hybrid objective / subjective test serves to prove that a defendant has not only acted unreasonably, but that they had the capacity to appreciate the nature and consequences of their criminal actions. Including the concept of volition, therefore, the thesis identifies criminal liability with the existence of three crucial mental capacities – the capacities to appreciate the nature and consequences of one's actions, to exercise

ordinary self-control, and to recognise and respond to reason. This, it is proposed, largely preserves the current understanding and application of legal responsibility within common law traditions, whilst resolving the contentions of unsafety and irrationality raised in the first part of the thesis.

The thesis presents some of the key implications of the resultant capacity-based theory of legal responsibility, for example, linking legal defences to deficiencies in one of more of the identified mental capacities, advocating for a new legal defence based on addiction, arguing against retributive theories of punishment and in favour of rehabilitative and deterrent theories, and offering a revised system of verdicts which may be rendered at the conclusion of a trial. Finally, the capacity-based theory of responsibility is generalised to govern responsibility for actions in general. This theory is reintroduced into the wider philosophical debate between free will and determinism where it is demonstrated how the theory provides a rational approach to holding people responsible for their actions in a deterministic universe and without any presumptions of, or reliance upon, metaphysical free will.

## *1.1. Research Question and Objectives*

The central research question in this thesis asks, *how can people rationally be held responsible for their actions in a deterministic universe absent of metaphysically free decision-making?* The thesis pursues three particular research objectives in order to answer this question:

1) To elucidate and expand upon theories of decision-making from neuroscience and psychology, and relate the current state of the art to relevant aspects of legal responsibility;
2) Drawing from the conclusions of the first objective, to appropriately reformulate the current conception of legal responsibility and *mens rea* in particular, taking into account the implications of current scientific research on decision-making, reasoning and volitional control; and

3) To place the theory within its broader philosophical background, suggesting key legal developments that are implied by the present research.

The thesis does not aim specifically to enter into the wider philosophical debate regarding the truth or falsity of the concepts of free will and determinism. Rather, the thesis takes the non-existence of free will and the truth of determinism as its basic underlying assumption. The overarching aim of the thesis, therefore, is to investigate the resultant implications of this assumption for the concept of legal responsibility and the method by which we hold some people criminally liable for their actions but others not.

Equally, the overarching aim of the thesis is not to completely undermine or replace existing methods of ascribing legal responsibility, which have proven to be effective and robust for centuries, not only in the UK but around the common law world. Rather, the aims of the thesis are meliorative; to take the present method of ascribing criminal liability and assess how it stands up against an alternative deterministic world view; and, in particular, to reform and redress the concept of legal responsibility in light of scientific research which provides an overwhelmingly deterministic description of how people arrive at (criminal) decisions.

## *1.2. Methodology and Structure*

The present thesis is completed using entirely secondary (desk-based) research methodologies. Part One of the thesis operates substantially as an extended literature review applying a conceptual research methodology. Thus, existing neuroscientific and psychological research concerning decision-making in the human brain is collated and analysed, drawing conclusions therefrom which have particular relevance to various aspects of the legal concept of responsibility, and specifically the component of *mens rea*. In Part Two of the thesis, the implications from the previous neuropsychological research are applied to the concept of legal responsibility in order to assess whether and to what extent this legal concept remains rational and defensible in light of science of human decision-making as it is currently understood. Part Two of the thesis further adopts the

meliorative aim of revising any such aspects of legal responsibility which are found to be irrational or incompatible with the previous scientific research.

A number of legal research methodologies are applied in Part Two of the thesis. A doctrinal methodology is used to present and describe the current approach to criminal responsibility as it exists in the English common law legal system. At various junctures, an historical-legal methodology as applied in order to illuminate the genesis and evolution of particular rules, principles and ideas within the broader concept of legal responsibilty. A socio-legal methodology is applied throughout Part Two of the thesis in order to explore the two-way relationship between society and the law – that is, to explain how behaviour and phenomena within society result in the development and evolution of the law and, indeed, how the law subsequently impacts upon behaviour and other phenomena within society. A normative "neuro-legal" methodology is further applied throughout Part Two of the thesis in order to analyse the compatibility of legal rules, principles, ideas and concepts against the conclusions drawn from the neuropsychological research explored in Part One of the thesis, and in order to affirm, amend, or entirely reform such legal concepts as is necessary and appropriate to achieve rational conformity between the law and neuropsychology.

Having arrived at a normative description of legal responsibilty which is rationally defensible against the neuropsychology of decision-making, Part Two of the thesis proceeds to analyse that theory of legal responsibilty by two principal means. First, a qualitative empirical analysis is conducted by applying the theory to leading UK jurisprudence (case law) in order to observe how the theory would apply in practice and to assess the hypothetical practical outcomes of the theory against the actual outcomes of individual cases. Second, the final substantive chapter of the thesis generalises the theory of *legal* responsibility into a theory of responsibility for human action *generally*. This general theory of responsibility is subjected to an analytic analysis to test its application to key arguments within the broader philosophical debate surrounding the compatibility of responsibility for action with causal determinism.

This plurality of research methods is justified by at least three aspects of the thesis. First, the subject matter of responsibility for action itself touches across numerous domains.

Aside from being a legal concept that is the main focus of the present research, responsibility is an important concept in moral philosophy; aspects of responsibility such as volition and intentionality are important points of investigation in their own right within neuroscience and psychology; and it is an important socio-political concept which has both developed and evolved over time and shaped society in the process. Second, and relatedly, the present thesis is an inherently multidisciplinary piece of research, beginning and ending with the fundamentally philosophical and metaphysical questions of compatibility between causal determinism and responsibility for action, whilst substantively drawing from the scientific study of neuropsychology and applying conclusions and implications therefrom to the legal and moral concepts of responsibilty. Such an inherently multidisciplinary investigation readily invites as range of different research methodologies.

The third justification relates to the underlying meliorative aims of the present thesis – that is, the intended purposes identifying and assessing shortcomings within current approaches to legal responsibility and, crucially, proposing amendments and reforms to redress those shortcomings in light of the neuropsychological research considered. To this end, the thesis is neither entirely descriptive nor entirely normative, but both; it is neither purely analytical nor purely dialectical, but both. Consequently, the range of research methodologies adopted are appropriate to fulfilling the multiple aims of drawing legal implications from scientific research, describing current approaches to legal responsibility, applying the former scientific implications to established legal principles and concepts, proposing theoretical and practical reforms to the law, and testing those proposals within empirical jurisprudence and theoretical philosophical discourse.

*

The thesis is completed with the following structure:

1) *Introduction* - provides the background and motivation behind the present thesis, setting out the research question and objectives, methodology and structure, justification for the research and underlying presumptions.

**Part One**

**2)** *The Expanded Brass-Haggard Model of Decision-Making* – presents and expands upon current models of decision-making from neuroscience and psychology. Brass and Haggard propose that any decision may be broken down into at least three components – the *what*, *when* and *whether* of a decision. The present thesis expands with the inclusion of *how* and *why* components, and further proposes how each component relates to the multi-alternative decision field theory of decision-making.

**3)** *The What Component, Priming and Predicting Choices* – considers first the *what* component of decision-making, with research from priming and fMRI studies suggesting the possibility for this component to operate automatically and without conscious effort or intervention, including in the formation of intentions and goals.

**4)** *The How Component and Sense of Agency* – explores the processes by which the brain plans how to carry out the physical actions required to enact a decision, and how this process of motor planning contributes to the overall sense of agency that people experience regarding their actions.

**5)** *The When Component and Timing of Consciousness* – presents the seminal experiments of Benjamin Libet and various repetitions and updates, together suggesting that the brain makes decisions prior to conscious awareness of those decisions.

**6)** *The Whether Component, Veto, and Impulse Control* – explores the ability to veto or cancel a decision that has already been prepared, again revealing unconscious neural correlates which suggest that controlling a decision, like making a decision in the first place, is a substantially unconscious process.

**7)** *The Why Component, Access to Reason and Post-hoc Rationalisation* – discusses the involvement of reasoning in the process of decision-making, suggesting that

reasons are constructed post-hoc in order to explain and justify a decision, rather than providing first principles from which the decision is arrived at.

**Part Two**

**8)** *Deconstructing Mens Rea* – this chapter presents some of the key implications from Part One of the thesis for the method of ascribing responsibility for criminal actions in the English common law tradition, critiquing in particular the law's reliance on proof of subjective mental states as a key determinant of legal responsibility.

**9)** *Reconstructing Mens Rea* – correspondingly, this chapter presents the key meliorative aspect of the thesis, replacing the concept of moral blame with "(un)reasonableness", and replacing the requirement for proof of specific subjective mental states with a hybrid objective / subjective test for *mens rea*.

**10)** *Elaborating Hybrid Objective / Subjective Mens Rea* – this chapter takes each of the different subjective mental states contained with the concept of *mens rea* – (*i.e.,* intention, recklessness, knowledge, *etc.*) – and demonstrates how each would be replaced within the hybrid objective / subjective concept. Further, this replacement concept is subsequently tested against key jurisprudence concerning each subjective mental state, demonstrating how the hybrid approach would apply in practice.

**11)** *Defences* – in a similar exercise to the preceding chapter, this chapter demonstrates how the various criminal defences can be rationalised within the modified approach to ascribing responsibility, in particular linking defences to the decision-making capacities which underpin the revised concept of *mens rea*. Further, a novel defence of addiction is presented and justified.

**12)** *Verdict and Punishment* – having presented and tested the revised, capacity-based concept of legal responsibility premised upon causal determinism and the denial

of metaphysical free will, the thesis proceeds to consider the implications of this concept of legal responsibility for theories of punishment, specifically admonishing retributive theories of punishment and advocating for deterrent and rehabilitative theories. The thesis further presents a novel verdict to modernise and replace the verdict of not guilty by reason of insanity.

**13)** *Philosophical Placement of the Present Thesis* – this final chapter places the present thesis within its broader philosophical context; it generalises the concept of *legal* responsibility into a capacity-based theory of moral responsibility and responsibility for actions generally; and it provides responses to various leading discussions in the philosophical debates surrounding free will, moral responsibility, and (in)compatibilism.

**14)** *Conclusions* – draws together from both Parts One and Two of the thesis to present the core conclusions to each research objectives, and answers the central research question.

## *1.3. Justification of Research and Original Contribution*

As indicated in the above introduction, modern legal systems around the world continue to ascribe responsibility for actions with approaches built upon the underlying presumption that human decisions and resultant (criminal) actions are "freely" chosen – that is, that agents could have chosen otherwise in the circumstances, and that they are the original authors of their decisions. Meanwhile, it is a generally held position amongst the natural sciences that the *macroscopic* universe is causally deterministic, and that all physical bodies larger than the atom are bound by the laws of causation – *i.e.,* cause and effect. This proposition applies no less to the human brain, and the disciplines of neuroscience and psychology largely reject any notion that the brain – a large, warm and wet physical object – could somehow operate outside of the fundamental laws of physics which otherwise govern the macroscopic universe.

Meanwhile, it is submitted that most (if not all) compatibilist approaches attempting to reconcile these philosophical positions are ultimately revisionist, accepting the incompatibilist rejection of metaphysical free will and instead redefining the concept in relation to some other criteria, such as the concurrence between first- and second-order desires or the existence of accessible reasons for decisions. Therefore, the underlying presumptions that the law holds regarding *metaphysical* free will, (upon which the concept of responsibility is premised), are generally opposed to the deterministic laws governing the natural sciences – not least those natural laws governing (macroscopic) physics, biology, chemistry and biochemistry – which underpin the functioning of the brain and the processes by which it reaches the very decisions for which people are held legally responsible.

The free will / determinism debate, of course, rages on within the fields of neuroscience and psychology, and neither study provides *conclusive* evidence in either direction – indeed, it is unlikely that either field of study is capable of definitively answering this fundamentally metaphysical question. Nonetheless, it is submitted that neuropsychology generally denies the *metaphysical* claims of free will – *i.e.*, the principle of alternative possibilities and the existence of uncaused causes. *Substance* dualism is widely refuted within these fields; the existence of a "Cartesian theatre" or homunculus in the brain that makes decisions independently of underlying brain activity is generally rejected. *Property* dualism obtains greater support in particular within psychology, with "weak" emergentism and epiphenomenalism remaining compatible with causal determinism (whereas any commitment to "strong" emergentism, like substance dualism, is widely considered as breaching the fundamental laws of physics). Meanwhile, monism and physicalism arguably obtain greater support within neuroscience, which is characterised by the goal of identifying the fundamental neural substrates from which human behaviour and experiences arise.

This broad disconnect between neuropsychological and legal presumptions regarding free will within a causally deterministic universe is fundamentally worthy of investigation if the entire concept of legal (and, indeed, moral) responsibility is found to rest upon false premises. In this regard, as indicated above, the present thesis is not *principally* intended

to enter into the broader philosophical debate regarding whether free will or determinism are true and / or compatible. Rather, the thesis goes one step beyond this in assuming the truth of determinism and the negation of free will as herein defined, and asks whether and to what extent the existing legal approaches remain a reliable, rational and fair means of attributing responsibility for actions to some people (*i.e.,* those who deserve to be so held responsible) whilst not attributing such responsibility to others (*i.e.,* those who, for whatever reason, do not deserve to be held responsible for their actions).

Moreover, whilst by no means determinative of the wider philosophical free will / determinism debate, the disciplines of neuroscience and psychology offer an increasingly detailed description of human decision-making and behaviour based significantly upon automatic, often unconscious, and ultimately deterministic processes. As all criminal conduct emanates from some (conscious or unconscious) decision to act (or not act), the scientific study of decision-making is exceptionally relevant to the legal question of responsibility for actions, making this topic eminently suited to the emerging study of "neurolaw". Meanwhile, the broadly deterministic description of decision-making processes offered by neuroscience and psychology present potential challenges to the concept of legal responsibility – and, indeed, related concepts such as punishment – where the latter legal concepts are built upon premises that may be fundamentally unsupported by the former scientific disciplines.

Neurolaw is a relatively modern discipline which, like the present thesis, seeks to investigate the implications of scientific research in the fields of neuroscience and psychology for various concepts and practices in the field of law. Such investigations could attempt to offer foundational principles drawn out from the scientific research and, using these as first principles, build atop legal concepts and practices which best align with the scientific state of the art. The reality, however, is that law is a far older discipline comprised of concepts that have been developed and ingrained over the course of centuries, if not millennia. It is unrealistic to expect that novel disciplines such as neurolaw would today completely overhaul and replace existing legal concepts and doctrines; that is to say, were a neuropsychological account of legal decision-making to be developed afresh by building from first principles, few jurisdictions around the world

will be prepared to replace entirely whole swathes of their legal system with novel, academic and unproven concepts developed from the new field of neurolaw. Rather, the study of neurolaw must be more modest in its ambitions, seeking to revise or amend legal concepts and practices according to the scientific research, rather than overhaul and replace those concepts and practices entirely. In this regard, the present thesis is justified by its strongly meliorative approach to the concept of legal responsibility – the thesis takes the currently existing method of ascribing liability for criminal acts from the English common law system and proposes how this should be revised and updated to better align with the scientific state of the art on human decision-making.

The present thesis makes original contributions in a number of areas. First, whereas delineating the topography of a decision into *what*, *how*, *when*, *whether* and *why* components is not of itself original, the thesis makes original contributions in the implications that are drawn from the scientific research in each of these areas to concepts within the legal topic of criminal responsibility. Second, the thesis makes original contributions in its revisions of the concept of *mens rea*; in particular, the development of the "reasonableness" principle to replace that of moral blameworthiness, and the development of a hybrid objective / subjective approach to *mens rea* replacing proof of subjective mental states are each original to the present thesis. Third, the thesis makes an original contribution in relating the various legal defences to three mental capacities which comprise the revised concept of *mens rea* in the capacity-based theory of responsibility developed in the thesis. Fourth and finally, the thesis offers original contributions to the wider philosophical debates surrounding free will, determinism and responsibility, and concerning (in)compatibilism between determinism and responsibility in particular.

# Contents

# PART ONE

# 2. The Expanded Brass-Haggard Model of Decision-Making

> '[E]ach action or event is "caused" by the cascade of events and influences that came before it, *ad infinitum*. Nothing originates with me. My thoughts and actions – my choice to pick up the pen or reach out to knock over the glass – are caused by electrical signals in my brain and the rest of my body, which in turn are caused by my particular body (including brain) and its reaction to this set of circumstances, which itself is the result of everything that I have ever experienced combined with the body I was born with, which developed from my genome, my various early environments, and their constant interaction, which in turn interact with the unfolding circumstances of my life. All of this was itself caused by a multitude of factors, and so on.'
>
> - David Wasserman and Josephine Johnston, 2014.[1]

Some human actions are the result of largely automatic processes – such as breathing and blinking – or purely biological reflexes – such as a knee-jerk response to striking the patellar tendon. All remaining behaviour and actions – and not least those actions to which we ascribe legal and moral responsibility – are otherwise the result of some decision made by the brain, which may be conscious or unconscious. The aggressive man who is knocked at a bar, spilling his drink, and who responds in the heat of the moment without due consideration by throwing a punch, does so as a result of some rapid, and often unconscious, cognitive process. Indeed, a recurrent theme throughout this thesis is that the vast majority of our decisions are in fact taken unconsciously, such as when we drive a familiar route to work and arrive with little recollection of the journey. Conversely, some of our decisions are the result of more consciously deliberative processes, such as

[1] David Wasserman and Josephine Johnston, 'Seeing responsibility: Can neuroimaging teach us anything about moral and legal responsibility' (2014) 44(s2) *Hastings Center Report* S37, S38.

the patient man in the same scenario, above, who realises the accident of having been knocked, and accepts an apologetic replacement drink rather than resorting to violence.

With those actions that are relevant to legal responsibility arising from conscious or unconscious decision-making processes, this chapter describes and expands upon a broad neuroscientific model of decision-making. First, the '*What, When, Whether Model of Intentional Action*'[2] by Marcel Brass and Patrick Haggard is presented, which rejects the notion of decision-making as a unitary concept and instead distinguishes three major components of any decision to act – 'deciding *what* to do, deciding *when* to do it, and deciding *whether* to implement one's decision or not.'[3] Second, this "Brass-Haggard model" is expanded upon by proposing two further components, namely *how* to execute a particular action, and *why* to do it. Third, the chapter integrates the expanded Brass-Haggard model with 'Decision Field Theory',[4] which ascribes the outcome of decision-making processes to the result of neuronal networks representing possible decision outcomes which compete to a threshold at which a particular decision is reached. Finally, the chapter discusses how the science of decision-making in the brain relates specifically to the way in which the law ascribes responsibility for some actions but not others.

## *2.1. The Brass-Haggard Model*

Brass and Haggard distinguish intentional action from reflex and purely stimulus-driven action, the former being both purposive and endogenous whilst the latter consists of an immediate stereotyped motor response caused by some external stimulus. Intentional actions do not *necessarily* possess an obvious external stimulus but may be initiated by wholly internal, endogenous states, and may be more 'flexible in form and timing, yet still be related to purpose.' Consequently, the human brain 'must make decisions or generate additional information, to produce intentional behaviours, which is not required for stimulus-driven action.' [5] Whereas neuroscience had previously attempted to

[2] Marcel Brass and Patrick Haggard, 'The what, when, whether model of intentional action' (2008) 14(4) *Neuroscientist* 319.
[3] *Ibid*., 320.
[4] Jerome R. Busemeyer and James T. Townsend, 'Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment' (1993) 100(3) *Psychological Review* 432.
[5] Brass and Haggard (2008), 319.

understand decision-making as a unitary concept, functional brain imaging research often portrayed a contradictory picture of the various different brain areas that are engaged in decision-making and intentional action. Rather than approaching intentional action as a unitary concept, Brass and Haggard propose at least three distinct components of any decision – *i.e.,* deciding *what* to do, *when* to do it, and *whether* or not to do it.

### *2.1.1. The "What" Component – Action Selection*

Brass and Haggard support their claim with reference to a number of experiments which disambiguate the involvement of different brain regions with each of the three claimed components for intentional action. The *what* component refers to action selection, *i.e.* the decision of what particular action to execute. Common experimental paradigms involve subjects making a free selection between available alternatives and contrasting the resultant brain activity with conditions where the subject's responses are triggered by external stimuli. Activity in various parts of the fronto-median wall appears indicative of free action selection, with the rostral cingulate zone ('RCZ') and the pre-supplementary motor area ('pre-SMA') being most consistently indicated.[6] However, one potentially conflating factor is the finding of similar fronto-median activity in relation to conflict resolution as opposed to action selection,[7] with activation potentially reflecting conflict between competing alternatives rather than the actual decision of what action to execute.[8]

In response, the authors highlight competition between different action alternatives as a crucial aspect of voluntary action itself – 'to decide for a specific behaviour, one has to

[6] Hakwan C. Lau, Robert D. Rogers, Narender Ramnani and Richard E. Passingham, 'Willed action and attention to the selection of action' (2004b) 21(4) *NeuroImage* 1407; Mark E. Walton, Joseph T. Devlin and Matthew F. S. Rushworth, 'Interactions between decision making and performance monitoring within prefrontal cortex' (2004) 7(11) *Nature Neuroscience* 1259; Veronika A. Mueller, Marcel Brass, Florian Waszak and Wolfgang Prinz, 'The role of the preSMA and the rostral cingulate zone in internally selected actions' (2007) 37(4) *NeuroImage* 1354.

[7] Matthew M. Botvinick, Todd S. Braver, Deanna M. Barch, Cameron S. Carter and Jonathan D. Cohen, 'Conflict monitoring and cognitive control' (2001) 108(3) *Psychological Review* 624; Parashkev Nachev, Geraint Rees, Andrew Parton, Christopher Kennard and Masud Husain, 'Volition and conflict in human medial frontal cortex' (2005) 15(2) *Current Biology* 122.

[8] K. Richard Ridderinkhof, Markus Ullsperger, Eveline A. Crone and Sander Nieuwenhuis, 'The role of the medial frontal cortex in cognitive control' (2004) 306(5695) *Science* 443; Matthew F. S. Rushworth, Mark E. Walton, Steven W. Kennnerley and David M. Bannerman, 'Action sets and decisions in the medial frontal cortex' (2004) 8(9) *Trends in Cognitive Sciences* 410.

overcome conflict from competing response alternatives.'[9] Moreover, the competition between alternatives is greater for entirely endogenous actions where, in the absence of external stimuli, the range of potential alternative actions have a more similar level of activation. Experimental manipulations attempting to dissociate action selection and response conflict have presented some contrasting results. For example, Lau *et. al.* (2006)[10] found greater activation in the RCZ in response to conflict between alternative actions and in the pre-SMA in response to intentional action selection; meanwhile, Nachev *et. al.*[11] found dissociation within the pre-SMA itself, with the more rostral area reflecting conflict resolution and the more caudal area reflecting free action selection.

However, Brass and Haggard suggest that this could be an artificial argument, 'because intentional action and response conflict may effectively be two sides of the same coin.'[12] The seminal psychologist William James writes, 'every mental representation of a movement awakens to some degree the actual movement which is its object; and awakens it in a maximum degree whenever it is not kept from so doing by an antagonistic representation present simultaneously in the mind.'[13] Expanding on this view, Brass and Haggard continue to suggest that 'response conflict is an inherent property of all action and intentional selection is necessarily required in such situations.'[14]

### *2.1.2. The "When" Component – Timing of a Decision*

The *when* component of a decision refers to the timing of intentional action; whereas, by definition, reflex actions are an immediate response to a stimulus, intentional actions occur at altogether more random times, and the timing of processes in the brain which result in intentional action has thus become an important area of research. Considerable attention has been given to the "readiness potential" (*bereitschaftspotential*) which is a

[9] Brass and Haggard (2008), 320; citing Botvinick *et. al.*; Parashkev Nachev, Henrietta Wydell, Kevin O'Neill, Masud Husain and Christopher Kennard, 'The role of the pre-supplementary motor area in the control of action' (2007) 36(3) *NeuroImage* T155.
[10] Hakwan C. Lau, Robert D. Rogers and Richard E. Passingham, 'Dissociating response selection and conflict in the medial frontal surface' (2006) 29(2) *NeuroImage* 446.
[11] Nachev *et. al.* (2005).
[12] Brass and Haggard (2008), 321.
[13] William James, *The Principles of Psychology* (MacMillan 1890), 1134.
[14] Brass and Haggard (2008), 321.

gradual increase in negativity in the motor areas of the brain which occurs a few seconds before intentional action. The absence of the readiness potential for actions resulting from instructional sensory cues provides particularly compelling evidence for its involvement specifically in the process of generating endogenous intentional actions.[15] A number of experiments suggest that the readiness potential begins first in the pre-SMA, followed by activation in motor areas of brain contralateral to the body part which gives effect to the action.[16]

As Brass and Haggard note, 'intentional actions are associated with an experience of endogenously initiating action' – the experience that '"I" control my actions.'[17] The most famous experimental paradigm is that by Libet *et. al.*[18] – discussed in detail in section 5.2 of this thesis – which investigated the timing of the readiness potential as compared with the timing of subjects' conscious experience of intending to act. Whereas a conscious intention to act was reported at an average of 206 milliseconds ('ms') prior to physically acting, the readiness potential was recorded much earlier from 1000 to 500ms prior to the onset of movement. With the readiness potential appearing to precede conscious awareness of an intention to act, it is reasoned that conscious thought itself cannot be the cause of the readiness potential but, rather, the readiness potential 'must cause both the movement and the conscious experience of being about to move.'[19] Although the paradigm has received notable criticism of both its methodology and interpretation, the results have been replicated in numerous subsequent studies.[20]

---

[15] Marjan Jahanshahi, Harri I. Jenkins, Richard G. Brown, David C. Marsden, Richard E. Passingham and David J. Brooks, 'Self-initiated versus externally triggered movements: I. An investigation using measurement of regional cerebral blood flow with PET and movement-related potentials in normal and Parkinson's disease subjects' (1995) 118(4) *Brain* 913.

[16] Tonio Ball, Axel Schreiber, Bernd Feige, Michael Wagner, Carl Hermann Lücking and Rumyana Kristeva-Feige, 'The role of higher-order motor areas in voluntary movement as revealed by high-resolution EEG and fMRI' (1999) 10(6) *NeuroImage* 682; Ross Cunnington, Christian Windischberger, Lüder Deecke and Ewald Moser, 'The preparation and readiness for voluntary movement: a high-field event-related fMRI study of the Bereitschafts-BOLD response' (2003) 20(1) *NeuroImage* 404.

[17] Brass and Haggard (2008), 321.

[18] Benjamin Libet, Curtis A. Gleason, Elwood W. Wright and Dennis Keith Pearl, 'Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act' (1983) 106(3) *Brain* 623.

[19] Brass and Haggard (2008), 322.

[20] For example, Patrick Haggard and Martin Eimer, 'On the relation between brain potentials and the awareness of voluntary movements' (1999) 126(1) *Experimental Brain Research* 128.

Further evidence for the distinct *when* component of decision-making is provided by Fried *et. al.*[21] – also discussed further in sections 3.2 and 5.2.4 of this thesis – in an experiment consisting of direct electrical stimulation to the supplementary motor area ('SMA'). Weaker electrical stimulation gave rise to the subjects reporting an urge to move particular body parts, whilst stimulation at higher levels resulted in actual movement, often of the same body part for which the urge was reported. As Brass and Haggard write, 'these results show that an experience that seems to be related to intention arises as part of the processes that lead to movement.'[22] Further experiments measuring the perceived timing of conscious intention to act in patients with focal parietal lesions also indicate the involvement of this area,[23] whilst reproductions of the Libet paradigm using function magnetic resonance imaging ('fMRI') have shown activation of the pre-SMA, intraparietal sulcus, and dorsolateral prefrontal cortex ('dlPFC') in relation to judging intention.[24] Brass and Haggard conclude that these experiments together suggest that 'intentional action is not generated by a single brain area but is a product of a recurrent fronto-parietal network.'[25]

### *2.1.3. The "Whether" Component – To Act or Not*

Before any decision is implemented through motor action, a final component of the decision-making process is to decide whether or not to initiate a particular decision into action. As Brass and Haggard write, 'in our daily life, we very often have to decide ourselves whether we should act or not' and, furthermore, 'overcoming impulsive behaviour is crucial for many cooperative social interactions.'[26] The authors cite their own earlier work[27] – an fMRI study broadly following the Libet paradigm in which

[21] Itzhak Fried, Amiram Katz, Gregory McCarthy, Kimberlee J. Sass, Peter Williamson, Susan S. Spencer and Dennis D. Spencer, 'Functional organization of human supplementary motor cortex studied by electrical stimulation' (1991) 11(11) *Journal of Neuroscience* 3656.
[22] Brass and Haggard (2008), 322.
[23] Angela Sirigu, Elena Daprati, Sophie Ciancia, Pascal Giraux, Norbert Nighoghossian, Andres Posada and Patrick Haggard, 'Altered awareness of voluntary action after damage to the parietal cortex' (2003) 7(1) *Nature Neuroscience* 80.
[24] Lau *et. al.* (2004b).
[25] Brass and Haggard (2008), 323.
[26] *Ibid*., 323.
[27] Marcel Brass and Patrick Haggard, 'To do or not to do: The neural signature of self-control' (2007) 27(34) *Journal of Neuroscience* 9141.

subjects were instructed to press a button at will whilst recording the timing of their subjective intention to make the motor action, except that they were told to sometimes (at their own choosing) inhibit that action at the last moment. Comparing brain activation when the motor action was carried through to completion with that when the action was inhibited, activity was found in the dorso-fronto-median cortex and anterior insula for the inhibition condition. A further study by Campbell-Meiklejohn *et. al.*[28] investigated brain activity in gamblers as they attempted to inhibit the strong behavioural tendency to continue gambling in order to try and recover losses, with the subjects again required to decide themselves whether or not to stop. The study similarly found activation in a region of the brain overlapping the dorso-fronto-median cortex.[29] Brass and Haggard submit that these data 'support the idea that the *whether* component of intentional action can be distinguished from the when and what components.'[30]

*

In their concluding remarks, Brass and Haggard suggest how the *what, when* and *whether* components of a decision may be further elucidated in a number of pathologies. Thus, it is proposed that anarchic hand syndrome might reflect an impairment of normal action selection and the *what* component; the difficulties of initiating movement for patients with Parkinson's disease may reflect an impairment of timing and the *when* component; and obsessive-compulsive disorders, Tourette's syndrome and attention deficit hyperactivity disorder ('ADHD') may reflect impairments in the *whether* component and the ability to inhibit actions.[31]

[28] Daniel K. Campbell-Meiklejohn, Mark W. Woolrich, Richard E. Passingham and Robert D. Rogers, 'Knowing when to stop: The brain mechanisms of chasing losses' (2008) 63(3) *Biological Psychiatry* 293.
[29] See also Adam R. Aron, Trevor W. Robbins and Russell A. Poldrack, 'Inhibition and the right inferior frontal cortex: One decade on' (2014) 18(4) *Trends in Cognitive Sciences* 177.
[30] Brass and Haggard (2008), 324.
[31] *Ibid*., 324.

### *2.1.4. Supporting Research*

#### *2.1.4.1. Theoretical and Meta-analytical Support*

Prior to the publication of Brass and Haggard's work, Jahanshahi and Frith[32] theorised a similar separation of volitional "willed" action into different components, at least comprising *what* to do, *when* to do it, and *whether* or not to do it. What is more, for each of these components, Jahanshahi and Frith equally identify a number of brain regions which appear to be engaged in particular in relation to volitional action as contrasted against externally triggered or stereotyped actions, including *inter alia* the dorsolateral prefrontal cortex, anterior cingulate cortex, and SMA.[33] Whilst this work was supported by a number of human and animal studies, some of these suffered from the relative infancy of neuroimaging techniques such as fMRI. Nevertheless, there is significant agreement on key brain areas being engaged between Jahanshahi and Frith's, and Brass and Haggard's work, with the separation of decision-making into at least three components first being theorised in the former paper, before receiving greater empirical justification in the latter work. Equally, the authors agree that deficiencies or pathologies within discrete circuits reflecting different components of decision-making may in turn manifest in different mental and physical illnesses such as Parkinson's disease and schizophrenia.

The most recent and significant review of the Brass-Haggard model is provided by Zapparoli, Seghezzi and Paulesu, who provide a meta-analytical review of studies published subsequent to the Brass-Haggard model.[34] Deploying hierarchical clustering and ALE meta-analytical procedures, Zapparoli, Seghezzi and Paulesu first identified a further 15 studies investigating the *what*, *when* or *whether* components, before searching through the BrainMap.org database for co-activations of identified brain regions.[35] The meta-analysis first confirms that a 'segregation of intention specific regions is possible even though the regions involved go beyond the mesial wall of the frontal lobe.'[36] This

[32] Marjan Jahanshahi and Christopher D. Frith, 'Willed action and its impairments' (1998) 15(6-8) *Cognitive Neuropsychology* 483.
[33] *Ibid*., 494.
[34] Laura Zapparoli, Silvia Seghezzi and Eraldo Paulesu, 'The what, the when, and the whether of intentional action in the brain: A meta-analytical review' (2017) 11 *Frontier in Human Neuroscience* 1.
[35] *Ibid*., 3 – 5.
[36] *Ibid*., 6.

partially confirms the Brass-Haggard model, suggesting a separation within the medial prefrontal cortex of the three components of decision-making, 'with the more anterior regions involved in more abstract decisions of whether to execute an action and the more posterior ones recruited in specifying the content and, yet more dorsally, the timing components of actions.'[37]

Regarding the *what* component, Zapparoli, Seghezzi and Paulesu found data clustered around the middle cingulum, which has previously been associated with the management and resolution of conflict between competing alternatives.[38] This finding lends support to Brass and Haggard's suggestion that action selection and conflict resolution may ultimately be two sides of the same coin; an idea tracing its roots back to William James. Regarding the *when* component, Zapparoli, Seghezzi and Paulesu identify a cluster in the SMA, an area previously associated with the timing or initiation of intentional movement.[39] This association is further bolstered by a number of studies of Parkinson's disease, which is characterised by an impaired ability to implement intentional actions and is widely hypothesised to result from a malfunctioning of the brain's mechanism for timing actions.[40] Finally, Zapparoli, Seghezzi and Paulesu identify a cluster in the anterior portion of the cingulum associated with the *whether* component, further supporting the suggestion that the decision of whether or not to act is separable from other aspects of volitional action.

Going beyond the dissociations in the median wall of the frontal lobe identified in common with Jahanshahi and Frith, and Brass and Haggard, the meta-analysis indicates further brain regions outside of this area which may also be engaged in decision-making

---

[37] *Ibid*., 7.

[38] *Ibid*., 7; citing Matthew M. Botvinick, Jonathan D. Cohen and Cameron S. Carter, 'Conflict monitoring and anterior cingulate cortex: an update' (2004) 8(12) *Trends in Cognitive Sciences* 539; Cameron S. Carter and Vincent van Veen, 'Anterior cingulate cortex and conflict detection: An update of theory and data' (2007) 7(4) *Cognitive, Affective, & Behavioural Neuroscience* 367.

[39] Zapparoli, Seghezzi and Paulesu (2017), 8; citing Cunnington, Windischberger, Deecke and Moser (2003); Filiep Debaere, Nichole Wenderoth, Stefan Sunaert, Paul van Hecke and Stephan P. Swinnen, 'Internal vs external generation of movements: differential neural pathways involved in bimanual coordination performed in the presence or absence of augmented visual feedback' (2003) 19(3) *NeuroImage* 764.

[40] Zapparoli, Seghezzi and Paulesu (2017), 8; citing Jahanshahi and Frith (1998); Brass and Haggard (2008); Jochen Michely, Lukas J. Volz, Michael T. Barbe, Felix Hoffstaedter, Shivakumar Viswanathan, Lars Timmermann, Simon B. Eickhoff, Gereon R. Fink and Christian Grefkes, 'Dopaminergic modulation of motor network dynamics in Parkinson's disease' (2015) 138(3) *Brain* 664.

in a component specific manner.[41] Again regarding the *what* component, data clustered around the supramarginal gyrus in the inferior parietal lobule, an area that has been shown to be critical for representing actions or an intention to act in previous studies.[42] Regarding the *when* component, clusters were found around the frontal operculum – an area previously indicted in the task of synchronising hand movements to an auditory rhythm[43] – and the lenticular nuclei, which are part of a cortical network regulating motor behaviour.[44] Finally, a number of clusters were found in relation to the *whether* component; first, the anterior insula is indicated, concurring with studies suggesting the involvement of this area in response inhibition[45] and concentration.[46] Second, the thalamus and putamen are indicated; these areas are previously known to play a role in action selection,[47] whilst their involvement in the inhibition of actions is suggested by their abnormal functioning in patients with Tourette's syndrome.[48]

[41] Zapparoli, Seghezzi and Paulesu (2017), 8.

[42] *Ibid*; citing Eugene Tunik, Nicola J. Rice, Antonia F. Hamilton and Scott T. Grafton, 'Beyond grasping: representation of action in human anterior intraparietal sulcus' (2007) 36(Supp. 2) *NeuroImage* T77; Jason P. Gallivan, D. Adam McLean, Kenneth F. Valyear, Charles E. Pettypiece and Jody C. Culham, 'Decoding action intentions from preparatory brain activity in human parieto-frontal networks' (2011) 31(26) *Journal of Neuroscience* 9599; Michel Desmurget, Karen T. Reilly, Nathalie Richard, Alexandru Szathmari, Carmine Mottolese and Angela Sirigu, 'Movement intention after parietal cortex stimulation in humans' (2009) 324(5928) *Science* 811.

[43] Michael H. Thaut, 'Neural basis of rhythmic timing networks in the human brain' (2003) 999(1) *Annals of the New York Academy of Sciences* 364.

[44] Ann M. Graybiel, 'The basal ganglia and chunking of action repertoires' (1998) 70(1-2) *Neurobiology of Learning and Memory* 119; Jill R. Crittenden and Ann M. Graybiel, 'Basal ganglia disorders associated with imbalances in the striatal striosome and matrix compartments' (2011) 5 *Frontiers in Neuroanatomy* 1.

[45] Tor D. Wager, Ching-Yune C. Sylvester, Steven C. Lacey, Derek Evan Nee, Michael Franklin and John Jonides, 'Common and unique components of response inhibition revealed by fMRI' (2005) 27(2) *NeuroImage* 323.

[46] Mark D. Allen, Erin D. Bigler, James Larson, Naomi J. Goodrich-Hunsaker and Ramona O. Hopkins, 'Functional neuroimaging evidence for high cognitive effort on the Word Memory Test in the absence of external incentives' (2007) 21(13-14) *Brain Injury* 1425.

[47] Mark D. Humphries and Kevin N. Gurney, 'The role of intra-thalamic and thalamocortical circuits in action selection' (2002) 13(1) *Network Computation in Neural Systems* 131; Mark D. Humphries, Robert D. Steward and Kevin N. Gurney, 'A physiologically plausible model of action selection and oscillatory activity in the basal ganglia' (2006) 26(50) *Journal of Neuroscience* 12921.

[48] Laura Zapparoli, Mauro Porta and Eraldo Paulesu, 'The anarchic brain in action: The contribution of task-based fMRI studies to the understanding of Gilles de la Tourette syndrome' (2015) 28(6) *Current Opinion in Neurology* 604.

#### *2.1.4.2. Recent Studies*

A number of studies subsequent to Brass and Haggard's 2008 paper have continued to provide evidence for separation of the decision-making process into at least these three components.[49] To begin, experiments have focused on further disentangling the *what* and *when* components of decision-making. Investigating the dividing line between these components, Krieghoff, Brass, Prinz and Waszak[50] developed a paradigm to explore both components within the same experiment. In brief, subjects had a choice of a button press and were cued at different times to decide which button to press (the *what* component) and *when* to do so. The subjects then heard four tones and had to press the chosen button on either the third or fourth tone, as previously decided. The separation of the timing of both the *what* and *when* decisions, and the later execution of the associated actions, allowed for these distinct decision components to be investigated through fMRI. The study observed different activation maxima for each component within two areas of the frontomedian wall, again supporting the dissociation of these decision-making elements.[51]

In concurrence with previous studies,[52] the RCZ showed the greatest activation in relation to the *what* component; however, in a departure from previous findings, a region of the superior frontal gyrus ('SFG') showed higher activity during the *when* component. The authors discuss the role of the RCZ in action selection and conflict monitoring as previously addressed in this chapter, above, and reach a similar conclusion as Willian James and Brass and Haggard that these two alternatives may in fact be complementary aspects of the same process.[53] The finding of activation in the left SFG in relation to the *when* component is 'to our knowledge the first evidence' for this particular association. The authors note that previous studies have instead indicated the involvement of the pre-SMA in the timing of intentional action, and they critique that former studies did not 'disentangle processes related to the decision when to act from processes related to the

[49] For a further review of supporting studies preceding Brass and Haggard (2008), see Veronika Krieghoff, Florian Waszak, Wolfgang Prinz and Marcel Brass, 'Neural and behavioral correlates of intentional actions' (2011) 49(5) *Neuropsychologia* 767, 772 – 774.
[50] Veronika Krieghoff, Marcel Brass, Wolfgang Prinz and Florian Waszak, 'Dissociating what and when of intentional actions' (2009) 3 *Frontiers in Human Neuroscience* 1.
[51] *Ibid*., 7.
[52] Citing Mueller, Brass, Waszak and Prinz (2007).
[53] Krieghoff, Bras, Prinz and Waszak, 7.

instantaneous initiation of the action and therefore presumably confounded these two factors.'[54] Moreover, the precise location of the activity recorded in the SFG was located very close to the pre-SMA, suggesting the probable existence of a functional link between the two areas.[55]

A further important finding by Krieghoff, Brass, Prinz and Waszak is that signal strength analysis suggests that the interaction of the *what* and *when* components within the paramedian frontal cortex indicates that these components are not entirely dissociated. They propose that this finding ought not to be surprising, however, given the interdependency of these components; 'for an action and its consequences to be evaluated both components have to be taken into account.'[56] They draw the analogy of a football player deciding whether to pass the ball or shoot for a goal, whereby the optimal choice depends upon when the player intends to act whilst the optimal timing equally depends upon the action being chosen. Nonetheless, the findings of, at least, partially dissociated processes for the *what* and *when* components continue to challenge the unitary account of decision-making and reveals 'voluntary action control [to be] an interplay of different neuroanatomically dissociable subfunctions.'[57]

Hoffstaedter, Grefkes, Zilles and Eickhoff[58] similarly investigate the *what* and *when* components; one aim of their study is to redress a perceived shortcoming in the work by Krieghoff *et. al.*, namely that the timing of the *when* component was limited to choosing between one of two tones and might therefore be regarded as a cued, rather than volitional, decision.[59] Deploying a button-press paradigm which permitted subjects to make an entirely self-timed decision as to *when* to act with use of fMRI, Hoffstaedter *et. al.* also find a dissociation between the *what* and *when* components. Specifically, main activation was recorded in the medial frontal cortex from the pre-SMA into the anterior midcingulate cortex ('aMCC') along with the bilateral dorsal premotor cortex ('dPMC') for the *what*

[54] *Ibid*., 7 – 8.
[55] *Ibid*., 8.
[56] *Ibid*.
[57] *Ibid*.
[58] Felix Hoffstaedter, Christian Grefkes, Karl Zilles and Simon B. Eickhoff, 'The "What" and "When" of self-initiated movement' (2012) 23(3) *Cerebral Cortex* 520.
[59] *Ibid*., 521.

component, and in the superior regions of the SMA and aMCC along with the anterior insula, putamen and globus pallidus for the *when* component.[60] This provides further evidence for a partial dissociation of the *what* and *when* components of decision-making.

Momennejad and Haynes[61] introduce an interesting additional dimension by investigating through fMRI the brain regions associated with the *what* and *when* components of *future*, as opposed to present, intentions. Subjects engaged in a paradigm requiring them to form an intention to be performed in the future, and were then engaged in a distractor task for a period of time before that future intention needed to be retrieved and executed. The results revealed the role of the anterior prefrontal cortex ('aPFC') in maintaining and retrieving future intentions. The information regarding the *what* component was encoded in the dorsomedial aPFC during maintenance and in the ventrolateral aPFC and lateral PFC during retrieval; meanwhile, information regarding the *when* component was encoded in the bilateral and medial aPFC during maintenance and in the dorsomedial aPFC and lateral PFC during retrieval.[62] Previous studies have suggested the involvement of the aPFC in encoding future goals[63] and prospective memory,[64] *i.e.* memory for planned future intention and actions. The findings therefore correlate with previous interpretations of the role of the aPFC, whilst the dissociation between *what* and *when* components continues to be demonstrated.

Focusing solely on the *what* component, Holroyd and Yeung[65] provide a review of research surrounding the anterior cingulate cortex ('ACC'), the RCZ having already been indicated in action selection and resolution of conflict between competing options.[66] The

[60] *Ibid*., 524 – 526.
[61] Ida Momennejad and John-Dylan Haynes, 'Human anterior prefrontal cortex encodes the "what" and "when" of future intentions' (2012) 61(1) *NeuroImage* 139.
[62] *Ibid*., 145.
[63] *Ibid*; citing Sylvain Charron and Etienne Koechlin, 'Divided representation of concurrent goals in the human frontal lobes' (2010) 328(5976) *Science* 360.
[64] *Ibid*; citing Paul W. Burgess, Sophie K. Scott and Christopher D. Frith, 'The role of the rostral frontal cortex (area 10) in prospective memory: a lateral versus medial dissociation' (2003) 41(8) *Neuropsychologia* 906; Craig P. McFarland and Elizabeth L. Glisky, 'Frontal lobe involvement in a task of time-based prospective memory' (2009) 47(7) *Neuropsychologia* 1660.
[65] Clay B. Holroyd and Nick Yeung, 'Motivation of extended behaviours by anterior cingulate cortex' (2012) 16(2) *Trends in Cognitive Sciences* 122.
[66] See this chapter, above.

disorder akinetic mutism, associated with lesions in the anterior midcingulate cortex,[67] manifests as a significant reduction in spontaneous speech and action despite the absence of any deficits in motor ability. This has led to a traditional association of the ACC with goal-directed behaviour.[68] Modern studies point towards a more specific role of the ACC in conflict monitoring between decision alternatives and cognitive control in decision-making,[69] upon which Holroyd and Yeung theorise that the ACC is responsible for selecting and maintaining decision options,[70] and 'learns to associate values with different options and chooses the appropriate option for the current environmental state.'[71]

Further, Holroyd and Yeung suggest that the ACC 'not only chooses the option but also determines the level of effort to be applied towards executing the policy, and maintains this signal until the option reaches its termination state.'[72] They propose that the ACC is supported in its function by the midbrain dopamine system, which provides "reward" for positive actions in the form of the neurotransmitter dopamine, and thus aids the ACC in learning the given value of different decision options. The engagement of the ACC with the dopamine system is particularly interesting from the perspective of conditions which distort or "highjack" the system and, in turn, impact significantly upon decision-making, most notably addiction disorders.

Having remarked on the *whether* component receiving the least experimental attention, Brass and Haggard worked in wider teams in 2009[73] and 2014[74] to investigate the neural correlates of inhibition specifically. Building upon their 2007 study, Kühn, Haggard and

[67] Brent A. Vogt, 'Regions and subregions of the cingulate cortex' in Vogt B. A. (ed.), *Cingulate Neurobiology and Disease* (Oxford University Press 2009), 23.
[68] Orrin Devinsky, Martha J. Morrell and Brent A. Vogt, 'Contributions of anterior cingulate cortex to behaviour' (1995) 118(1) *Brain* 279, 296 – 297.
[69] For review, see Nicholas Yeung, 'Conflict monitoring and cognitive control' in Ochsner K. N. and Kosslyn S. (eds.), *The Oxford Handbook of Cognitive Neuroscience: Volume 2: The Cutting Edges* (Oxford University Press 2013).
[70] Holroyd and Yeung (2012), 123.
[71] *Ibid*., 125.
[72] *Ibid*.
[73] Simone Kühn, Patrick Haggard and Marcel Brass, 'Intentional inhibition: How the "veto-area" exerts control' (2009) 30(9) *Human Brain Mapping* 2834.
[74] Margot A. Schel, Simone Kühn, Marcel Brass, Patrick Haggard, K. Richard Ridderinkhof and Eveline A. Crone, 'Neural correlates of intentional and stimulus-driven inhibition: a comparison' (2014) 8 *Frontiers in Human Neuroscience* 1.

Brass developed a computer-based paradigm where subjects, recorded by fMRI, must press a button to prevent a green marble from rolling down a ramp and smashing, or must otherwise prepare this action but then may choose whether or not to inhibit it if the marble is white, thus exploring the un-cued and endogenous inhibition of action. Subjects reported the position of the marble on the ramp at the time when they reached the decision to inhibit or not, thus providing a measurement of the subjective timing of this decision.

The results indicated that the RCZ was active when deciding between acting and inhibiting, which reflects a decision of the *what* component between two alternatives. However, the dorsal fronto-median cortex ('dFMC') was only active in the process of self-initiated inhibition of pre-planned actions, *i.e.,* the *whether* component,[75] concurring with Brass and Haggard's 2007 work. Going further, Kühn, Haggard and Brass reveal functional connectivity between the dFMC and pre-SMA, the latter of which is indicated in action selection and conflict resolution and, furthermore, *may* provide the genesis of the readiness potential. The authors suggest that this connectivity 'provides an intentional mechanism for stopping an ongoing action in a top-down fashion' whereby 'inputs from dFMC to pre-SMA therefore potentially control whether actions occur or not.'[76]

In the 2014 study, Schel, Kühn, Brass, Haggard, Ridderinkhof and Crone investigate the neural correlates of inhibition through comparing intentional and stimulus-driven inhibition through the aforementioned marble paradigm alongside a classic stop-signal task.[77] Through a side-by-side comparison of the tasks, the authors determined that both stimulus-driven and intentional inhibition recruited similar networks in the right inferior frontal gyrus ('IFG') and pre-SMA – regions more traditionally associated with intentionality[78] – however, further activation in the bilateral inferior parietal lobe ('IPL') and pre-SMA in relation specifically to intentional inhibition 'suggest[s] that the

[75] Kühn, Haggard and Brass (2009), 2841.
[76] *Ibid*., 2842.
[77] See Gordon D. Logan and Willian B. Cowan, 'On the ability to inhibit thought and action: A theory of an act of control' (1984) 91(3) *Psychological Review* 295.
[78] Hakwan C. Lau, Robert D. Rogers, Patrick Haggard and Richard E. Passingham, 'Attention to intention' (2004a) 303(5661) *Science* 1208; Thilo van Eimeren, Thomas Wolbers, Alexander Münchau, Christian Büchel, Cornelius Weiller and Hartwig Roman Siebner, 'Implementation of visuospatial cues in response selection' (2006) 29(1) *NeuroImage* 286.

inhibition process cannot be reduced to intentionality *per se*.'[79] Furthermore, in concurrence with previous results, activity recorded in the dFMC was demonstrated to be sensitive to the demands of the particular tasks engaged in by subjects, for example, where the task induced a 'prepotency of responding' as opposed to inhibiting a response.[80]

### *2.1.5. Summary of the Brass-Haggard Model*

In a subsequent work, which also provides a broad overview of evidence supporting the Brass-Haggard model, Brass, Lynn, Demanet and Rigoni[81] offer a comprehensive summary of the theory:

> 'Our model assumes that early stages of intentional action are related to anterior prefrontal brain regions. These brain regions process complex and heterogenous information that is only broadly determined by specific task instruction or goals. Processing in these brain regions provides a sort of informational background, or intuition, and has a biasing function towards later processing stages. This complex set of information is funnelled when information travels more posteriorly and enters later stages of intentional action. Regions in the RCZ are related to choices between different response options. Such choices are biased by bottom-up information but also by concrete instructions that operate as a top-down influence and thus are a result of the interplay between top-down and bottom-up processing. Furthermore, the RCZ determines the level of effort that is invested in pursuing a specific behaviour and thus regulates the "willpower" that is invested in a specific choice. When a specific response option is selected, this information is transferred to brain areas more closely related to the motor system, namely SMA/pre-SMA. Here, the impulse to initiate a specific response is generated. At this point in the processing stream, it is still possible to disengage from the intention to act or to change the

[79] Schel, Kühn, Brass, Haggard, Ridderinkhof and Crone, 9.
[80] *Ibid*.
[81] Marcel Brass, Margaret T. Lynn, Jelle Demanet and Davide Rigoni, 'Imaging volition: What the brain can tell us about the will' (2013) 229(3) *Experimental Brain Research* 301.

intended behaviour. Intentional inhibition is achieved by a signal from the dorsomedial prefrontal cortex that downregulates activation in the SMA/pre-SMA. As a working hypothesis, we assume that the subjective experience of volition results from supra-threshold activation in brain circuits that are involved in the control of intentional action. Such subjective experiences are phenomenologically rich because they can be related to any level of the processing stream, ranging from intuitive feelings to concrete urges to act.'[82]

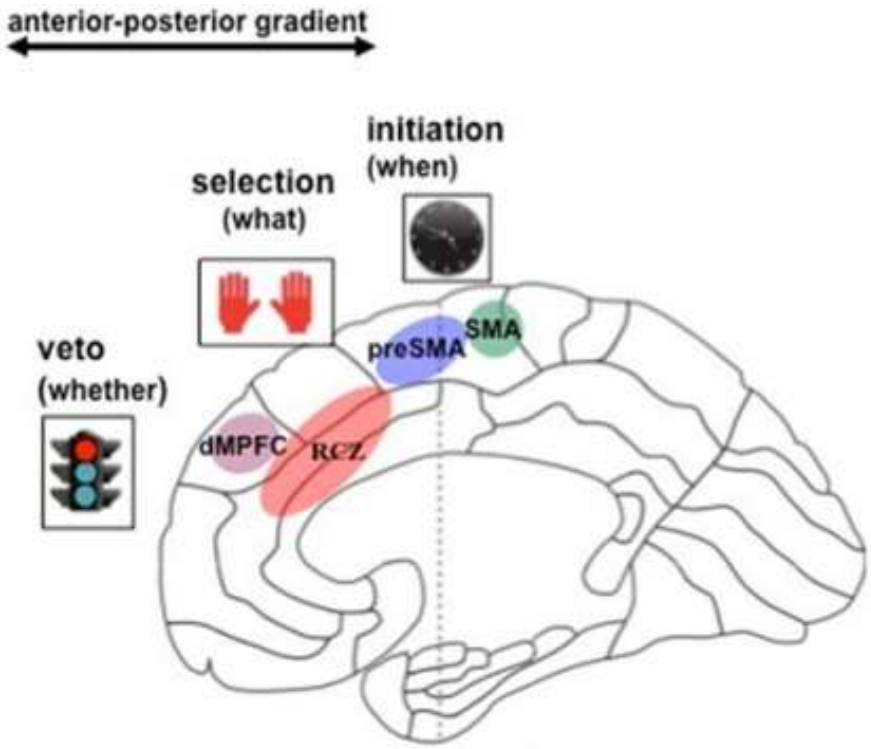


***Fig. a – The* what, when, whether *model of decision-making and related brain regions.*[83]**

## 2.2. Expanding the Model

The described Brass-Haggard model posits the, at least partial, dissociation of *what*, *when* and *whether* components of decision-making and volitional action; this thesis proposes that the model may be expanded to include two further components, namely the *how* and *why* components. The *how* component is considered first as, similarly to the components

[82] *Ibid.*, 309.
[83] *Ibid.*, 303.

considered thus far, this component is a necessary antecedent to initiating a volitional action. Conversely, whereas a degree of human decisions are undoubtedly goal-driven, many other decisions are not so, and the relation of a decision with a specific goal (*i.e., why* to do something) is not a *necessary* antecedent to reaching a decision or initiating an action. As is discussed further in chapter seven of this thesis, below, the reasons for why people reach certain decisions or engage in particular actions may often be confabulated by the brain, with reasons for decisions largely being constructed *post hoc* after the decision itself has already been taken.

### *2.2.1. The "How" Component*

Pre-dating the work of both Brass and Haggard, and Jahanshahi and Frith, the earlier work of Deecke and Kornhuber[84] theorised a separation of volitional action into three, slightly varied, components; *what* to do, *how* to do, and *when* to do. Regarding the *how* component, Deecke suggests that the frontolateral cortex is primarily responsible for deciding how to carry out a particular action, highlighting strong connections between this region and the sensory association areas of the parietal lobes. He writes, 'quick decisions regarding the tactics of "how" (*i.e.,* what is the best way) to achieve the goal requires always the newest information about the sensory situation.'[85]

Exploring the neural correlates of planning and execution to inhibit continuing actions, Omata, Ito, Takata and Ouchi[86] begin by noting previous research that has indicated the involvement of the pre-SMA and SMA,[87] premotor cortex and IPL[88] in the planning of

[84] Lüder Deecke, 'Planning, preparation, execution, and imagery of volitional action' (1996) 3(2) *Cognitive Brain Research* 59.
[85] *Ibid*., 60.
[86] Kei Omata, Shigeru Ito, Youhei Takata and Yasuomi Ouchi, 'Similar neural correlates of planning and execution to inhibit continuing actions' (2018) 12 *Frontiers in Neuroscience* 1.
[87] Ross Cunnington, Christian Windischberger and Ewald Moser, 'Premovement activity of the pre-supplementary motor area and the readiness for action: Studies of time-resolved event-related functional MRI' (2005) 24(5-6) *Human Movements Science* 644; Hiroshi Shibasaki, 'Cortical activities associated with voluntary movements and involuntary movements' (2012) 123(2) *Clinical Neurophysiology* 229.
[88] Desmurget, Reilly, Richard, Szathmari, Mottolese and Sirigu (2009); Silmar Teixeira, Sergio Machado, Bruna Velasques, Antonio Sanfim, Daniel Minc, Caroline Peressutti, Juliana Bittencourt, Henning Budde, Mauricio Cagy, Renato Anghinah, Luis F. Basile, Roberto Piedade, Pedro Ribeiro, Cláudia Diniz, Consuelo Cartier, Mariana Gongora, Farmy Silva, Fernanda Manaia and Julio Guilherme Silva, 'Integrative parietal cortex processes: Neurological and psychiatric aspects' (2014) 338(1-2) *Journal of the Neurological Sciences* 12.

volitional actions. The authors present a paradigm in which subjects reproduced a finger-tapping rhythm and, at certain junctures, were required to plan but subsequently inhibit the action, using fMRI to investigate brain activation during planning, execution and inhibition of movements. For the planning of action – which relates to the *how* component – the study recorded activation in the SMA and pre-SMA, the IPL, the IFG, the dlPFC, the insula, the left cerebellum, the primary visual cortex, and the globus pallidus / putamen. The authors add that the SMA, MCC, IPL, IFG and insula were similarly activated during the execution phase of inhibiting a planned action, whilst activation in the dlPFC, globus pallidus / putamen, visual cortex and cerebellum was particular to the planning phase only.[89] The authors further note previous associations of the dlPFC with complex executive functions including making plans for the future;[90] the globus pallidus with the regulation of movement;[91] the striatum and putamen with associating actions and rewards and selecting between competing alternatives;[92] and the cerebellum in the coordination, planning and execution of movement and motor control.[93]

Anderson and Cui[94] provide an overview of recent research – including a number of primate studies – indicating several roles of the strongly interconnected posterior parietal and frontal cortical areas in decision-making and action planning in particular. At the more abstract level of planning decisions or future actions, the IPL and parietal reach region ('PRR') have been found to be engaged in encoding the expected reward value of potential movements arising out of a decision to act,[95] whilst 'neurons in the putamen and

[89] Omata, Ito, Takata and Ouchi (2018), 8.
[90] Sam J. Gilbert and Paul W. Burgess, 'Executive function' (2008) 18(3) *Current Biology* R110.
[91] Rita Moretti and Riccardo Signori, 'Neural correlates for apathy: Frontal-prefrontal and parietal cortical-subcortical circuits' (2016) 8 *Frontiers in Aging Neuroscience* 1.
[92] Mimi Liljeholm and John P. O'Doherty, 'Contributions of the striatum to learning, motivation, and performance: an associative account' (2012) 16(9) *Trends in Cognitive Sciences* 467.
[93] Sarah-Jayne Blakemore and Angela Sirigu, 'Action prediction in the cerebellum and in the parietal lobe' (2003) 153(2) *Experimental Brain Research* 239; Jeremy D. Schmahmann, 'The role of the cerebellum in cognition and emotion: Personal reflection since 1982 on the dysmetria of thought hypothesis, and its historical evolution from theory to therapy' (2010) 20(3) *Neuropsychology Review* 236.
[94] Richard A. Anderson and He Cui, 'Intention, action planning, and decision making in parietal-frontal circuits' (2009) 63(5) *Neuron* 568.
[95] *Ibid*., 572; citing Michael L. Platt and Paul W. Glimcher, 'Neural correlates of decision variables in parietal cortex' (1999) 400(6741) *Nature* 233; Leo P. Sugrue, Greg S. Corrado and William T. Newsome, 'Matching behavior and the representation of value in the parietal cortex' (2004) 304(5678) *Science* 1782; Tianming Yang and Michael N. Shadlen, 'Probabilistic reasoning by neurons' (2007) 447(7148) *Nature* 1075.

caudate nucleus have been shown to encode action value.'[96] At the level of direct action preparation, active involvement of the posterior parietal cortex ('PPC') is indicated in the planning of motor movements,[97] with discrete areas of the parietal cortex even being dissociable to particular types of movements in primate experiments. For example, the anterior intraparietal area ('AIP') both 'appears selective for grasps' [98] and is 'interconnected with the ventral premotor cortex'[99] ('PMv') which 'also has activity related to grasp movements.'[100] Furthermore, neurons in the IPL of primates have been shown to encode both specific motor acts and the observed acts of others, offering a gateway through which an observer can understand the intentions of an observed agent.[101] Electrical stimulation of the IPL in humans has been found to induce a 'strong intention and desire to move their body parts', further suggesting a role for the PPC in awareness of intentionality.[102]

One recent 2015 experimental paradigm by Ariani, Wurm and Lingnau isolates the movement planning component of decision-making and uses fMRI to search for areas of activation across internally- and externally-driven plans to move.[103] In particular, subjects were either instructed, or had a free choice, to plan a different precision, power or touching motion. Three key results were obtained: first, activity in the superior parietal lobule ('SPL'), intraparietal sulcus ('IPS'), dorsal premotor cortex ('PMd') and primary

[96] *Ibid*; citing Kazuyuki Samejima, Yasumasa Ueda, Kenji Doya and Minoru Kimura, 'Representation of action-specific reward values in the striatum' (2005) 310(5752) *Science* 1337.
[97] *Ibid*., 568; citing John F. Kalaska, Stephen H. Scott, Paul Cisek and Lauren E Sergio, 'Cortical control of reaching movements' (1997) 7(6) *Neurobiology* 849.
[98] *Ibid*., 568; citing Hideo Sakata, Masato Taira, Makoto Kusunoki, Akira Murata and Yuichiro Tanaka, 'The TINS Lecture: The parietal association cortex in depth perception and visual control of hand action' (1997) 20(8) *Trends in Neuroscience* 350; Markus A. Baumann, Marie-Christine Fluet and Hansjörg Scherberger, 'Context-specific grasp movement representation in the macaque anterior intraparietal area' (2009) 29(20) *Journal of Neuroscience* 6434.
[99] *Ibid*; citing Judith Tanné-Gariépy, Eric M. Rouiller and Driss Boussaoud, 'Parietal inputs to dorsal versus ventral premotor areas in the macaque monkey: evidence for largely segregated visuomotor pathways' (2002) 145(1) *Experimental Brain Research* 91.
[100] *Ibid*; citing Giacomo Rizzolatti, Rosolino Camarda, Leonardo Fogassi, Maurizio Gentilucci, Giuseppe Luppino and Massimo Matelli, 'Functional organization of inferior area 6 in the macaque monkey' (1988) 71(3) *Experimental Brain Research* 491.
[101] See Leonardo Fogassi, Pier Francesco Ferrari, Benno Gesierich, Stefano Rozzi, Fabien Chersi and Giacomo Rizzolatti, 'Parietal lobe: From action organization to intention understanding' (2005) 308(5722) *Science* 662.
[102] Anderson and Cui (2009), 569; citing Desmurget, Reilly, Richard, Szathmari, Mottolese and Sirigu (2009).
[103] Giacomo Ariani, Mortiz F. Wurm and Angelika Lingnau, 'Decoding internally and externally driven movement plans' (2015) 35(42) *Journal of Neuroscience* 14160.

motor cortex ('M1') contralateral to the hand being moved was found to be associated with action planning regardless of whether that action was internally or externally cued. Second, activity was found in the contralateral ventral premotor cortex ('PMv'), dlPFC and supramarginal gyrus ('SMG'), and ipsilateral posterior IPS, posterior superior temporal gyrus ('STG') and posterior middle temporal gyrus ('MTG') for internally-, but not externally-, driven movement planning. Third, activity was recorded in the bilateral SMA and pre-SMA for encoding externally-driven movement plans.

Marneweck and Flamand[104] interpret the first collection of findings in the SPL, IPS, PMd and M1 as representing 'movement plans that were invariant to the way they were selected' – *i.e.,* the *how* component of a decision – whilst suggesting that the second collection of findings in PMv, dlPFC, SMG, IPS, STG and MTG likely represent the choice of *what* action to perform out of the three different motor actions that could be chosen in the study.[105] As they write, the work by Ariani, Wurm and Lingnau further demonstrates 'the involvement of neuroanatomically dissociable regions for different decision components in generating voluntary actions within the fronto-median wall.'[106] That being said, the results also suggest a degree of closeness or interconnectedness between the *what* and *how* components, with similar brain regions being engaged when subjects were engaged in internally-driven movement planning, possibly representing the selection (*i.e., what*) between various options plans (*i.e., how*). This potential connection between the *what* and *how* components is further elucidated in section 4.1 of this thesis, below.

Finally, Ptak, Schnider and Fellrath provide an overview of the most recent research, framing the dorsal frontoparietal network ('dFPM') in general as providing a 'core system for emulated action.'[107] They propose that the dFPM 'evolved as an extension of a simple action-control network connecting the posterior parietal cortex… with the PMd, and that the cognitive functions of this network are rooted within its fundamental capacity to

[104] Michelle Marneweck and Véronique H. Flamand, 'Elucidating the neural circuitry underlying planning of internally-guided voluntary action' (2016) 116(6) *Journal of Neurophysiology* 2469.
[105] *Ibid*., 2470.
[106] *Ibid*.
[107] Redek Ptak, Armin Schnider and Julia Fellrath, 'The dorsal frontoparietal network: A core system for emulated action' (2017) 21(8) *Trends in Cognitive Sciences* 589.

support planning and imagining actions.'[108] Thus, the dFPM network is responsible for providing an 'abstract, offline representation of movements expressed in terms of their pragmatics (action goals) and kinematics (movement patterns).'[109] The existence of such an emulation process is 'inferred from the observation that motor control relies on predictions of the consequences of motor plans' (see ***figure b***).[110]

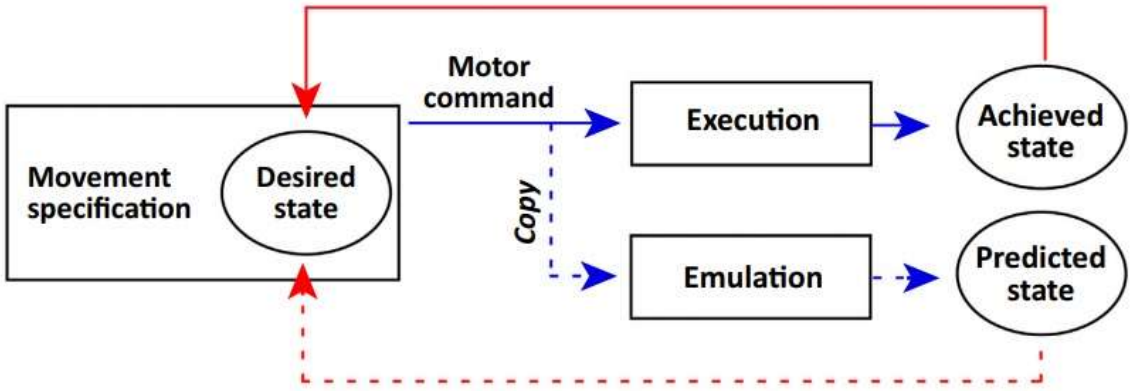


***Fig. b – Motor control relies on predictions of the consequences of motor plans, produced through emulation processes in the dFPM network.***[111]

Ptak, Schnider and Fellrath reason that the online adaptation and control required for performing skilled actions relies upon sensory feedback provided continuously throughout a given motor movement.[112] The utility of feedback processes for predicting the future outcome of actions is limited,[113] however, and models of action control instead propose that 'motor planning entails the anticipation of the predicted state of an effector compared with its desired state.'[114] This is achieved through a 'forward model of the motor-to-sensory transformation required for successful action'[115] within which an 'emulator receives input from processes involved in motor planning and computes a

[108] *Ibid*., 589.
[109] *Ibid*., 590.
[110] *Ibid*.
[111] Ptak, Schnider and Fellrath (2017), 590.
[112] *Ibid*; citing Aaron L. Wong, Adrian M. Haith and John W. Krakauer, 'Motor planning' (2015) 21(4) *Neuroscientist* 385.
[113] *Ibid*; citing Rachael D. Seidler, Douglas C. Noll and G. Thiers, 'Feedforward and feedback processes in motor control' (2004) 22(4) *NeuroImage* 1775.
[114] *Ibid*; citing Daniel M. Wolpert and Zoubin Ghahramani, 'Computational principles of movement neuroscience' (2000) 3(supp) *Nature Neuroscience* 1212.
[115] *Ibid*; citing David W. Franklin and Daniel M. Wolpert, 'Computational mechanisms of sensorimotor control' (2011) 72(3) *Neuron* 425; Michel Desmurget and Angela Sirigu, 'A parietal-premotor network for movement intention and motor awareness' (2009) 13(10) *Trends in Cognitive Science* 411.

forward model of proprioceptive and kinematic output just as though the movement had been performed.'[116] Ptak, Schnider and Fellrath draw evidence for this model from studies revealing invariant brain activity across the frontoparietal cortex across imagined versus performed movements,[117] endogenously initiated versus externally triggered movements,[118] and right- versus left-hand movements.[119]

One of the central tenets behind Ptak, Schnider and Fellrath's action emulation account is that, having regard to the dFPM being a common substrate for both motor and cognitive processes, a 'more complex cognitive process may emerge from simpler processes if it shares neural sources and uses overlapping computational mechanisms.'[120] Various studies are cited as providing direct evidence for this assertion; for example, the superior parietal cortex, IPS and precuneus have all been indicated in relation to executing simple arm movements such as pointing and reaching, hand movements such a grasping, and coordinated finger movements such as are engaged in writing and drawing.[121] These same activities have equally been demonstrated to elicit activity in the PMd.[122] Further, activity recorded in the PMd, SPL and IPS has been shown to be a robust predictor of various forms of motor learning.[123] Moreover, the dFPN is activated to a similar degree in reach- and grasp-related activities, whether the particular movement is performed with the arm visible or in darkness,[124] whether or not movements are delayed (thus engaging working

[116] *Ibid*; citing Rick Grush, 'The emulation theory of representation: motor control, imagery, and perception' (2004) 27(3) *Behavioral and Brain Sciences* 377.
[117] *Ibid*; Nikolaas N. Oosterhof, Steven P. Tipper and Paul E. Downing, 'Visuo-motor imagery of specific manual actions: a multi-variate pattern analysis fMRI study' (2012) 63(1) *NeuroImage* 262.
[118] *Ibid*; citing Ariana, Wurm and Lingnau (2015).
[119] *Ibid*; Jason P. Gallivan, D. Adam McLean, J. Randall Flanagan and Jody C. Culham, 'Where one hand meets the other: limb-specific and action-dependent movement plans decoded from preparatory signals in single human frontoparietal brain areas' (2013) 33(5) *Journal of Neuroscience* 1991.
[120] *Ibid*; 591.
[121] *Ibid*., 592; citing Alexandra Battaglia-Mayer, Lucy Babicola and Eleonora Satta, 'Parieto-frontal gradients and domains underlying eye and hand operations in the action space' (2016) 334 *Neuroscience* 76; Guy Vingerhoets, 'Contribution of the posterior parietal cortex in reaching, grasping, and using objects and tools' (2014) 5 *Frontiers in Psychology* 151; Flavia Filimon, Jonathan D. Nelson, Ruey-Song Huang and Martin I. Soreno, 'Multiple parietal reach regions in humans: Cortical representations for visual and proprioceptive feedback during on-line reaching' (2009) 29(9) *Journal of Neuroscience* 2961.
[122] *Ibid*; citing Flavia Filimon, 'Human cortical control of hand movements: parietofrontal networks for reaching, grasping, and pointing' (2010) 16(4) *Neuroscientist* 388.
[123] *Ibid*., 592; citing Robert M. Hardwick, Claudia Rottschy, R. Chris Miall and Simon B. Eickholl, 'A quantitative meta-analysis and review of motor learning in the human brain' (2013) 67 *NeuroImage* 283.
[124] *Ibid*; citing Filimon, Nelson, Huand and Soreno (2009).

memory),[125] and when movements are simply observed.[126] Ptak, Schnider and Fellrath conclude that an 'overwhelming number of neuroimaging studies indicate that motor planning and imagery,' amongst other functions, engage the dFPN, confirming the association of this network with the *how* component of decision-making.

### *2.2.2. The "Why" Component*

Arguably the hallmark of *volitional* action is that decisions and their resultant actions can typically be understood as pursuing some goal, purpose or reason which underlies and motivates a particular decision to act. Disambiguating this *why* component to single brain regions or networks is difficult, perhaps because the very notion of what constitutes the goal of a decision can encompass many things. For example, ***figure c*** (below) envisages the concept of a goal in two parts: a broader "task" goal such as preparing food; and a number of discrete "immediate goals" which must be accomplished in order to achieve the task goal, such as taking a cookie and pouring a glass of milk.

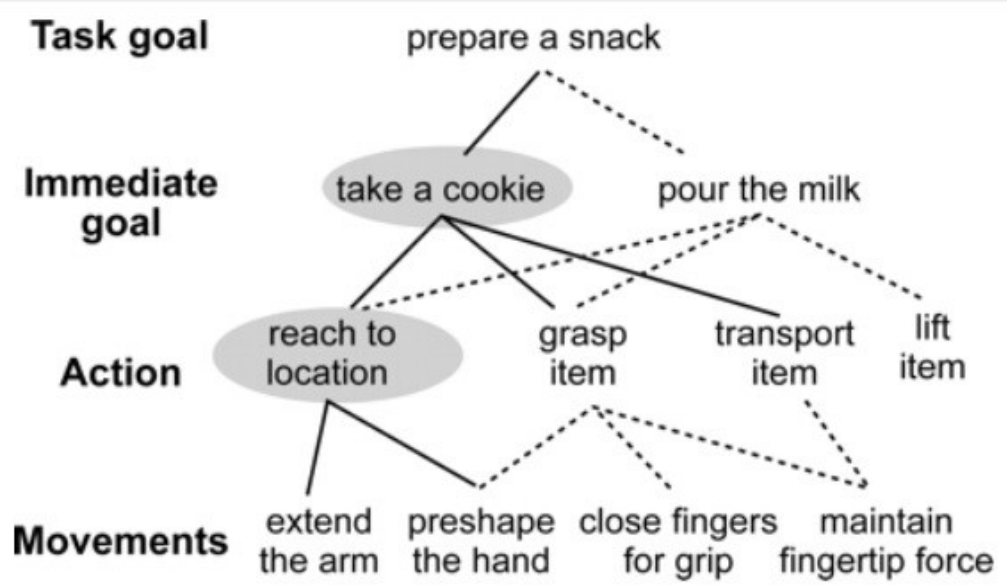


***Fig. c – The hierarchical organisation of goals.***[127]

[125] *Ibid*; citing Katja Fiehler, Michael M. Bannert, Matthias Bischoff, Carlo Blecker, Rudolf Stark, Dieter Vaitl, Volker H. Franz and Frank Rösler, 'Working memory maintenance of grasp-target information in the human posterior parietal cortex' (2010) 54(3) *NeuroImage* 2401.

[126] *Ibid*; citing Svenja Caspers, Karl Zilles, Angela R. Laird and Simon B. Eickhoff, 'ALE meta-analysis of action observation and imitation in the human brain' (2010) 50(3) *NeuroImage* 1148.

[127] Antonia F. de C. Hamilton and Scott T. Grafton, 'Goal representation in human anterior intraparietal sulcus' (2006) 26(4) *Journal of Neuroscience* 1133, 1134.

Exploring a similar type of model which disambiguates goal-directed choice into *why, what, where, when* and *how* components, Verschure, Penartz and Pezzulo identify several processes that potentially contribute to the formation of goals which underly decisions.[128] At the most fundamental level, the task goal behind a given decision or action may relate to the various basic biological needs of an individual person or, indeed, animal. In this sense, decisions to get or make food satisfy the goal of hunger, and retrieving a drink satisfies thirst; putting on or taking off clothes may satisfy the feeling of being too cold or hot; going to sleep satisfies tiredness, *etc.* For these purposes, 'both the sensor and effector functions of the hypothalamus are critical.'[129] The hypothalamus – alongside other lower central nervous system centres including the brain stem, spinal cord and autonomic ganglia – is responsible for monitoring, generating and dissipating body heat;[130] the osmolality of blood plasma and homeostasis of salt levels;[131] nutrients and energy;[132] sleep and arousal;[133] and sexual and maternal behaviours.[134] As Verschure, Penartz and Pezzulo write, the hypothalamus and other lower-order structures give rise to basic drives and their associated behavioural expressions such as hunger, aggression and sleep. A drive 'arises from the discrepancy between a read-out of a homeostatic parameter (*e.g.,* blood sugar level) and an optimal set point, although for some types of drives the neural basis underlying this comparison is not that clear yet.'[135]

Verschure, Penartz and Pezzulo continue to propose that, once the 'needs of an agent ("Why") have been set at the level of the hypothalamus and brain stem, representations

[128] Paul F. M. J. Verschure, Cyriel M. A. Pennartz and Giovanni Pezzulo, 'The why, what, where, when and how of goal-directed choice: neuronal and computational principles' (2014) 369(1655) *Philosophical Transactions of the Royal Society: Biological Sciences* 20130483.
[129] *Ibid*., 20130488.
[130] Shaun F. Morrison, Kazuhiro Nakamura and Christopher J. Madden, 'Central control of thermogenesis in mammals' (2008) 93(7) *Experimental Physiology* 773.
[131] Charles W. Bourque, 'Central mechanisms of osmosensation and systemic osmoregulation' (2008) 9(7) *Neuroscience* 519.
[132] Clémence Blouet and Gary J. Schwartz, 'Hypothalamic nutrient sensing in the control of energy homeostasis' (2010) 209(1) *Behavioural Brain Research* 1.
[133] J. Gregor Sutcliffe and Luis de Lecea, 'The hypocretins: Setting the arousal threshold' (2002) 3(5) *Nature Reviews Neuroscience* 339.
[134] Loretta M. Flanagan-Cato, 'Sex differences in the neural circuit that mediates female sexual receptivity' (2011) 32(2) *Frontiers in Neuroendocrinology* 124; Danielle S. Stolzenberg and Michael Numan, 'Hypothalamic interaction with the mesolimbic DA system in the control of the maternal and sexual behaviors in rats' (2011) 35(3) *Neuroscience & Biobehavioral Reviews* 826.
[135] Verschure, Penartz and Pezzulo (2014), 20130488 – 20130489.

of the state of the world (including the agent's own state) are required to determine where and when this need may be satisfied, and through which particular object ("What") within a feasible spatio-temporal range.'[136] The hippocampus in particular has been associated with both representing the state of the world and the agent's place within it, 'incorporating many types of causal and / or non-causal spatio-temporal relationships', and actively storing and retrieving episodic memories and other relevant information.[137] The authors update two classical views of the hippocampal system as it relates to the formation of goal-directed behaviour. On the one hand, the hippocampus is traditionally associated with encoding an agent's position in space;[138] this view is updated with a large body of research indicating that the hippocampus encodes not only for the agent's position in space but also for other specific objects and events.[139]Moreover, the 'representation of the task seems to follow a multiplexing of input streams combining sensory, location and action information at both the input and memory states of hippocampal processing.'[140]

On the other hand, the hippocampus is also traditionally associated with the recording of experience in the form of episodic memories to be transferred to other neocortical areas where the generalisation (or 'semanticization') of memory occurs.[141] Verschure, Penartz and Pezzulo update this view, adding that the hippocampus also – 'and more generally' –

[136] *Ibid*., 20130489.
[137] *Ibid*., 20130490.
[138] Edward C. Tolman, 'Cognitive maps in rats and men' (1948) 55(4) *Psychological Review* 189; John O'Keefe and Jonathan Dostrovsky, 'The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat' (1971) 34(1) *Brain Research* 171.
[139] Emma R. Wood, Paul A. Dudchenko and Howard Eichenbaum, 'The global record of memory in hippocampal neuronal activity' (1999) 397(6720) *Nature* 613; Stefan Leutgeb, Jill K. Leutgeb, Carol A. Barnes, Edvard I. Moser, Bruce L. McNaughton and May-Britt Moser, 'Independent codes for spatial and episodic memory in hippocampal neuronal ensembles' (2005) 309(5734) *Science* 619; Carien S. Lansink, Jadin C. Jackson, Jan V. Lankelma, Rutsuko Ito, Trevor W. Robbins, Barry J. Everitt and Cyriel M. A. Pennartz, 'Reward cues in space: Commonalities and differences in neural coding by hippocampal and ventral striatal ensembles' (2012) 32(36) *Journal of Neuroscience* 12444; Benjamin J. Kraus, Robert J. Robinson II, John A. White, Howard Eichenbaum and Michael E. Hasselmo, 'Hippocampal "time cells": Time versus path integration' (2013) 78(6) *Neuron* 1090.
[140] Verschure, Penartz and Pezzulo (2014), 20130489; citing Robert U. Muller and John L. Kubie, 'The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells' (1987) 7(7) *Journal of Neuroscience* 1951; César Rennó-Costa, John E. Lisman and Paul F. M. J. Verschure, 'The mechanism of rate remapping in the dentate gyrus' (2010) 68(6) *Neuron* 1051; César Rennó-Costa, John E. Lisman and Paul F. M. J. Verschure, 'A signature of attractor dynamics in the CA3 region of the hippocampus' (2014) 10(5) *PLoS Computational Biology* e1003641.
[141] Verschure, Penartz and Pezzulo (2014), 20130489; citing Endel Tulving, 'Episodic memory: From mind to brain' (2002) 53(1) *Annual Review of Psychology* 1.

records 'chains of associated events and sequences of motor actions.'[142] Moreover, the hippocampus is subsequently able to retrieve information that it has previously stored, as well as 'self-generate internal sequences of cell activity that are subsequently used to map novel environments or situations.'[143] Thus, as Verschure, Penartz and Pezzulo explain, the hippocampus plays various roles both during ongoing goal-directed behaviour and when "off-line", having 'multiple modes to (re)generate and recall information from memory, which can be flexibly used to guide decision-making and / or support consolidation.'[144]

Verschure, Penartz and Pezzulo highlight further evidence for the role of the prefrontal cortex in goal-directed behaviour and, more specifically, in representing a 'task space, *i.e.,* the set of rules, constraints, goals and goal-predictive values of cues and actions available as options to pursue goals.'[145] For example, the authors cite studies in non-human animals indicating the role of prefrontal neurons in encoding the requisite task rules that must be followed to achieve an end goal;[146] particular actions or groups thereof

[142] *Ibid*., 20130489; citing Howard Eichenbaum, Paul Dudchenko, Emma Wood, Matthew Shapiro and Heikki Tanila, 'The hippocampus, memory, and place cells: Is it spatial memory or a memory space?' (1999) 23(2) *Neuron* 209; Laure Rondi-Reig, Géraldine H. Petit, Christine Tobin, Susumu Tonegawa, Jean Mariani and Alain Berthoz, 'Impaired sequential egocentric and allocentric memories in forebrain-specific-NMDA receptor knock-out mice during a new task dissociating strategies of navigation' (2006) 26(15) *Neuroscience* 4071; Henrique O. Cabral, Martin Vinck, Celine Fouquet, Cyriel M. A. Pennartz, Laure Rondi-Reig and Francesco P. Battaglia, 'Oscillatory dynamics and place field maps reflect hippocampal ensemble processing of sequence and place memory under NMDA receptor control' (2014) 81(2) *Neuron* 402.

[143] *Ibid*., 20130490; citing Matthew A. Wilson and Bruce L. McNaughton, 'Reactivation of hippocampal ensemble memories during sleep' (1994) 265(5172) *Science* 676; George Dragori and Susumu Tonegawa, 'Preplay of future place cell sequences by hippocampal cellular assemblies' (2010) 469(7330) *Nature* 397; Margaret F. Carr, Shantanu P. Jadhav and Loren M. Frank, 'Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval' (2011) 14(2) *Nature Neuroscience* 147.

[144] *Ibid*; citing Brad E. Pfeiffer and David J. Foster, 'Hippocampal place-cell sequences depict future paths to remembered goals' (2013) 497(7447) *Nature* 74; Giovanni Pezzulo, Matthijs A. A. van der Meer, Carien S. Lansink and Cyriel M. A. Pennartz, 'Internally generated sequences in learning and executing goal-directed behavior' (2014) 18(12) *Trends in Cognitive Sciences* 647.

[145] *Ibid*; citing Frédérique Kouneiher, Sylvain Charron and Etienne Koechlin, 'Motivation and cognitive control in the human prefrontal cortex' (2009) 12(7) *Nature Neuroscience* 939.

[146] Jonathan D. Wallis, Kathleen C. Anderson and Earl K. Miller, 'Single neurons in prefrontal cortex encode abstract rules' (2001) 411(6840) *Nature* 953.

which lead up to a goal;[147] and the representation of goals and goal sites themselves.[148] With regards to this latter function in particular, neurons in the orbito-frontal and medial prefrontal-anterior cingulate have been shown to be 'sensitive to the motivational value of cues,'[149] as well as 'actions associated with goal pursuit.'[150] In addition, lesion studies in animals further indicate towards the causal role of prefrontal structures in the representation of goals and task rules,[151] and learning the relationships between actions and outcomes.[152]

Further considering the role of the PFC in representing decision goals, Gazzaniga, Ivry and Mangun first describe an anterior-posterior gradient across the cortex that varies according to levels of abstraction.[153] Therefore, more abstract representations involve the more anterior regions of the PFC, whilst less abstract representations involve more posterior regions; 'in the extreme, we might think of the most posterior part of the frontal lobe, the primary motor cortex, as the point where abstract intentions are translated into concrete movement.'[154] Thus, a similar hierarchical organisation is reasoned in relation

[147] Bruno B. Averbeck, Jeong-Woo Sohn and Daeyeol Lee, 'Activity in prefrontal cortex during dynamic selection of action sequences' (2006) 9(2) *Nature Neuroscience* 276; Mark H. Histed and Earl K. Miller, 'Microstimulation of frontal cortex can reorder a remembered spatial sequence' (2006) 4(5) *PLoS Biology* e134.

[148] Vincent Hok, E. Save, Pierre-Pascal Lenck-Santini and Bruno Poucet, 'Coding for spatial goals in the prelimbic/infralimbic area of the rat frontal cortex' (2005) 102(12) *Proceedings of the National Academy of Sciences* 4602; Satoshi Tsujimoto, Aldo Genovesio and Steven P. Wise, 'Transient neuronal correlations underlying goal selection and maintenance in prefrontal cortex' (2008) 18(12) *Cerebral Cortex* 2748.

[149] Verschure, Penartz and Pezzulo, 20130490; citing Léon Tremblay and Wolfram Schultz, 'Relative reward preference in primate orbitofrontal cortex' (1999) 398(6729) *Nature* 704; Geoffrey Schoenbaum, Barry Setlow, Michael P. Saddoris and Michael Gallagher, 'Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala' (2003) 39(5) *Neuron* 855; Camilo Padoa-Schioppa and John A. Assad, 'Neurons in the orbitofrontal cortex encode economic value' (2006) 441(7090) *Nature* 223.

[150] *Ibid*; citing Steven W. Kennerley, Aspandiar F. Dahmubed, Antonio H. Lara and Jonathan D. Wallis, 'Neurons in the frontal lobe encode the value of multiple decision variables' (2009) 21(6) *Journal of Cognitive Neuroscience* 1162; Takayuki Hosokawa, Steven W. Kennerley, Jennifer Sloan and Jonathan D. Wallis, 'Single-neuron mechanisms underlying cost-benefit analysis in frontal cortex' (2013) 33(44) *Journal of Neuroscience* 17385.

[151] Rebecca Dias, Trevor W. Robbins, Angela C. Roberts, 'Dissociation in prefrontal cortex of affective and attentional shifts' (1996) 380(6569) *Nature* 69; Jennifer M. Birrell and Verity J. Brown, 'Medial frontal cortex mediates perceptual attentional set shifting in the rat' (2000) 20(11) *Journal of Neuroscience* 4320.

[152] Bernard W. Balleine, A. Simon Killcross and Anthony Dickinson, 'The effect of lesions of the basolateral amygdala on instrumental conditioning' (2003) 23(2) *Journal of Neuroscience* 666.

[153] Michael S. Gazzaniga, Richard B. Ivry and George R. Mangun, *Cognitive Neuroscience: The Biology of the Mind* (5th ed. W. W. Norton & Co. 2019), 525; citing Randall O'Reilly, 'The what and how of prefrontal cortical organization' (2010) 3(8) *Trends in Neurosciences* 355.

[154] *Ibid*., 525.

to decision goals, with more abstract goals – such as long-term career plans – represented more anteriorly, and more concrete goals – such as applying for a particular job – represented more posteriorly, until reaching the motor areas responsible for the physical movements necessary for the goal of filling out a particular job application form.[155]

Gazzaniga, Ivry and Mangun cite two experiments which provide particular support for this hierarchical arrangement of goal representation in the PFC. In the first experiment by Badre and D'Esposito,[156] subjects in an fMRI study were presented with tasks of increasing complexity; the simplest task related a number of available finger responses according to the colour of different squares presented one by one (Level A); an additional level of complexity was introduced in which the finger response related to a texture, which was in turn related to the coloured squared (Level B); further complexity still was introduced whereby the coloured squares must be used to determine upon which dimension to judge whether two stimuli matched (Level C); and finally a similar variation whereby the mapping of colours or dimensions was varied between blocks (Level D).

Consistent with the hypothesis of an anterior-posterior gradient across the PFC, activation remained in the most posterior, premotor regions for the simplest tasks, whereas increasingly anterior regions were recruited to achieve more complex and abstracted goals. A later replication of the experiment by Badre, Hoffman, Cooney and D'Esposito[157] used subjects with various focal frontal lobe lesions, and found that subjects with the most anterior lesions performed similar to controls on the easier Level A and B tasks but worse on the more complex Level C and D tasks. Meanwhile, subjects with more posterior lesions over the premotor cortex were impaired across all of the tasks. This evidence supports the hierarchical nature of the anterior-posterior gradient across the PFC.

[155] *Ibid*., 539.
[156] David Badre and Mark D'Esposito, 'Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex' (2007) 19(12) *Journal of Cognitive Neuroscience* 2082.
[157] David Badre, Joshua Hoffman, Jeffrey W. Cooney and Mark D'Esposito, 'Hierarchical cognitive control deficits following damage to the human frontal lobe' (2009) 12(4) *Nature Neuroscience* 515.

Finally, a further fMRI study by Hamilton and Grafton[158] explored the representation of immediate goals, which are 'characterised by the conjunction of a particular object with a particular action sequence, for example, reaching, grasping, and taking a cookie.'[159] Subjects were shown various short video clips depicting a hand making reaching, grasping and taking movements towards one of two objects reflecting different goals, whilst observing for brain activity differing between actions expressing novel or repeated goals. The results indicated that 'repeated observation of an action directed toward the same goal results in a systematic reduction of activation in the left intraparietal sulcus', suggesting this region in particular as being involved in the representation of immediate goals from the observed actions of others.[160] This finding follows the theory of repetition suppression, which suggests that the repetition of a stimulus may result in the suppression of blood oxygen level-dependent signals in the relevant regions that code that stimulus,[161] whilst the experimental design manipulates different video clips to isolate novel and familiar goals expressed through different actions. It must be noted, however, that repetition suppression has not previously been used for studying motor representations in this way.

Specifically, Hamilton and Grafton found immediate goals represented in 'two regions of the lateral bank of IPS, within the inferior parietal lobe.'[162] These areas have previously been shown to activate in response to observing hand actions,[163] whilst damage to the same is shown to impair the ability for people to interpret the actions of others.[164] Moreover, the inferior parietal cortex is regarded as being part of the 'human mirror system',[165] with more recent evidence demonstrating the coding of objects in the IPS,

[158] Hamilton and Grafton (2006).
[159] *Ibid*., 1133.
[160] *Ibid*., 1135.
[161] *Ibid*., 1133; citing Kalanit Grill-Spector and Rafael Malach, 'fMRI-adaptation: a tool for studying the functional properties of human cortical neurons' (2001) 107(1-3) *Acta Psychologica* 293; Lionel Naccache and Stanislas Dehaene, 'The priming method: Imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes' (2001) 11(10) *Cerebral Cortex* 966.
[162] Hamilton and Grafton (2006), 1136.
[163] Julie Grèzes and Jean Decety, 'Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis' (2000) 12(1) *Human Brain Mapping* 1.
[164] Leslie J. Gonzalez Rothi, Kenneth M. Heilman and Robert T. Watson, 'Pantomime comprehension and ideomotor apraxia' (1985) 48(3) *Neurosurgery & Psychiatry* 207.
[165] Hamilton and Grafton (2006), 1136; citing Giacomo Rizzolatti and Laila Craighero, 'The mirror-neuron system' (2004) 27(1) *Annual Review of Neuroscience* 169.

'possibly related to a goal representation.'[166] Activity reflecting representations in the more anterior IPS cluster further overlaps with a region previously associated with grasping actions,[167] and the correction of grasping actions to pursue a new goal is delayed when this same area is disrupted using transcranial magnetic stimulation ('TMS').[168] This suggests that the IPS 'maintains a representation of the current goal to correct for errors' during action, whilst the evidence taken together indicates that the parietal cortex is a 'critical region for the representation of actions plans and goals.'[169] This latter point in particular is further reflected in the discussion at section 2.2.1 and chapter four of this thesis, where the IPS was discussed to be similarly involved in planning discrete movements under the *how* component. It may be theorised that the same area is engaged in maintaining a representation of the immediate goal and preparing the motor actions necessary – *i.e., how* – to complete that goal, whilst correcting for errors as those actions are carried out.

## 2.3. Competing Neural Networks

The previous sections 2.1 and 2.2 have disambiguated decision-making into five different components, representing the *what*, *how, when*, *whether* and *why* of any given individual decision, and have further explored evidence for the separate representation of these different components across, at least partially, distinct brain regions and networks. However, this does not explain how these various different regions, or the brain as a whole, actually reaches a decision between two or more competing options. Indeed, this question is itself unsettled; the present section of the thesis therefore provides a brief overview of the leading theories seeking to address this question, and considers how these theories might be linked with the expanded Brass-Haggard model, discussed above.

[166] *Ibid*; citing Lior Shmuelof and Ehud Zohary, 'Dissociation between ventral and dorsal fMRI activation during object and action recognition' (2005) 47(3) *Neuron* 457.
[167] Scott H. Frey, Deborah Vinton, Roger Norlund and Scott T. Grafton, 'Cortical topography of human anterior intraparietal cortex active during visually guided grasping' (2005) 23(2-3) *Cognitive Brain Research* 397.
[168] Eugene Tunik, Scott H. Frey and Scott T. Grafton, 'Virtual lesions of the anterior intraparietal area disrupt goal-dependent on-line adjustments of grasp' (2005) 8(4) *Nature Neuroscience* 505.
[169] Hamilton and Grafton (2006), 1136.

### *2.3.1. Multi-alternative Decision Field Theory*

"Decision Field Theory" was first presented by Busemeyer and Townsend as a 'mathematical foundation leading to a dynamic, stochastic theory of decision behaviour in an uncertain environment.'[170] Whereas the original theory considered decisions between two options, the authors later expanded the theory to incorporate the various relationships between choice, selling prices and certainty equivalents and,[171] later, expanded the theory further still to 'accommodate multi-alternative preferential choice situations.'[172] As the authors describe, the 'basic intuition underlying decision field theory is that a decision maker's preference for each option evolves during deliberation by integrating a stream of comparisons of evaluations among options on attributes over time.'[173] Thus, groups of neurons representing different decision options gather "support" or "evidence" towards each option or, equally, are suppressed from a lack of such support and competition from other options, until a threshold is reached which represents the arrival at a final decision. "Multi-alternative Decision Field Theory" has proven to be powerful in explaining a number of important findings arising from preferential choice studies such as the similarity effect,[174] attraction effect,[175] and compromise effect.[176]

Take, for example, the decision of which new car to purchase between A, B and C; initially, attention might be focused on a single most important attribute (such as quality) along with some particular aspects of that attribute (such as acceleration, control, stability and braking distance). This attribute and its aspects are evaluated for a period of time during which each purchase option is 'compared with others and these comparisons

[170] Busemeyer and Townsend (1993), 432.
[171] James T. Townsend and Jerome Busemeyer, 'Dynamic representation of decision-making' in Port R. F. and van Gelder T. (eds.), *Mind as Motion: Explorations in the Dynamics of Cognition* (Massachusetts Institute of Technology 1995).
[172] Robert M. Roe, Jerome R. Busemeyer and James T. Townsend, 'Multialternative decision field theory: A dynamic connectionist model of decision making' (2001) 108(2) *Psychological Review* 370.
[173] *Ibid*., 372.
[174] See Lennart Sjöberg, 'Choice frequency and similarity' (1977) 18(1) *Scandinavian Journal of Psychology* 103; Amos Tversky, 'Elimination by aspects: A theory of choice' (1972) 79(4) *Psychological Review* 281.
[175] See Srinivasan Ratneshwar, Allan D. Shocker and David W. Stewart, 'Toward understanding the attraction effect: The implications of product stimulus meaningfulness and familiarity' (1987) 13(4) *Journal of Consumer Research* 520; Itamar Simonson, 'Choice based on reasons: The case of attraction and compromise effects' (1989) 16(2) *Journal of Consumer Research* 158.
[176] See Simonson (1989); Itamar Simonson and Amos Tversky, 'Choice in context: Tradeoff contrast and extremeness aversion' (1992) 29(3) *Journal of Marketing Research* 281; Amos Tversky and Itamar Simonson, 'Context-dependent preferences' (1993) 39(10) *Management Science* 1179.

change the preferences up or down depending on whether an option has an advantage or disadvantage on the attended attribute.'[177] Attention later switches to a different and less crucial attribute (such as economy) and, again, comparisons of specific related aspects (such as price, fuel efficiency, reliability and durability *etc.*) are combined with the previous preferences. Such comparisons may thus continue between different attributes and aspects until, eventually, 'a decision is reached either by an externally imposed time constraint (*e.g.,* the car dealer presses for a final decision) or by a self-imposed criterion (*e.g.,* preference exceeds a threshold and the buyer announces a decision).'[178]

Multi-alternative decision field theory falls within a larger set of decision models known as sequential sampling models, which incorporate a number of core elements. In brief, each available alternative to a decision possesses an associated "valence" value representing the 'momentary advantage or disadvantage of option *i* when compared with other options on some attribute under consideration.'[179] A "valence vector" contains an ordered set of valences for all available options and is determined by three components: the 'personal evaluation of each option on each attribute', the 'attention weight allocated to each attribute at a particular moment in time', and the 'comparison process that contrasts the weighted evaluations of each option.'[180] Further, each available alternative to a decision possesses a "preference strength" representing the 'integration of all the valences considered for alternative *i* up to that point in time.'[181] New preference states are updated following an equation which provides a 'weighted combination of the previous preference state and the new input valence', whilst the overall dynamic behaviour of the model is determined by the initial preference state at the beginning of the decision and a feedback matrix.[182]

Ultimately, the evolving preference states determine the final outcome of a decision, according to one or more of a number of stopping rules. Thus, in the decision between purchasing cars A, B or C, above, the decision time may be imposed externally by the

[177] Roe, Busemeyer and Townsend (2001), 372.
[178] *Ibid*.
[179] *Ibid*., 372 – 373.
[180] *Ibid*., 373.
[181] *Ibid*., 373 – 374.
[182] *Ibid*., 374.

dealer losing patience and pressing for a decision, or internally once the strength of a particular preference crosses a given threshold. In ***figure d***, below, the output preference state for each of the three cars is represented by the three trajectories as they evolve over time. The vertical line to the left represents an externally controlled stopping rule, at which point the decision time is forced and car B is chosen, having the highest preference state at that particular point in time. The horizontal line to the top represents an internally controlled stopping rule, at which point a threshold has been reached by car A which is, therefore, the outcome of the decision.[183] In ***figure e***, below, the three trajectories similarly represent preference states between five cars A to E, the vertical lines represent shifting attention between different attributes, and the horizontal line at the bottom represents a boundary to discard. Thus, as different attributes are considered, some of the available options are discarded because they do not suit the decision-maker's preferences, and the stopping rule may consist of waiting for the last option to survive being rejected.

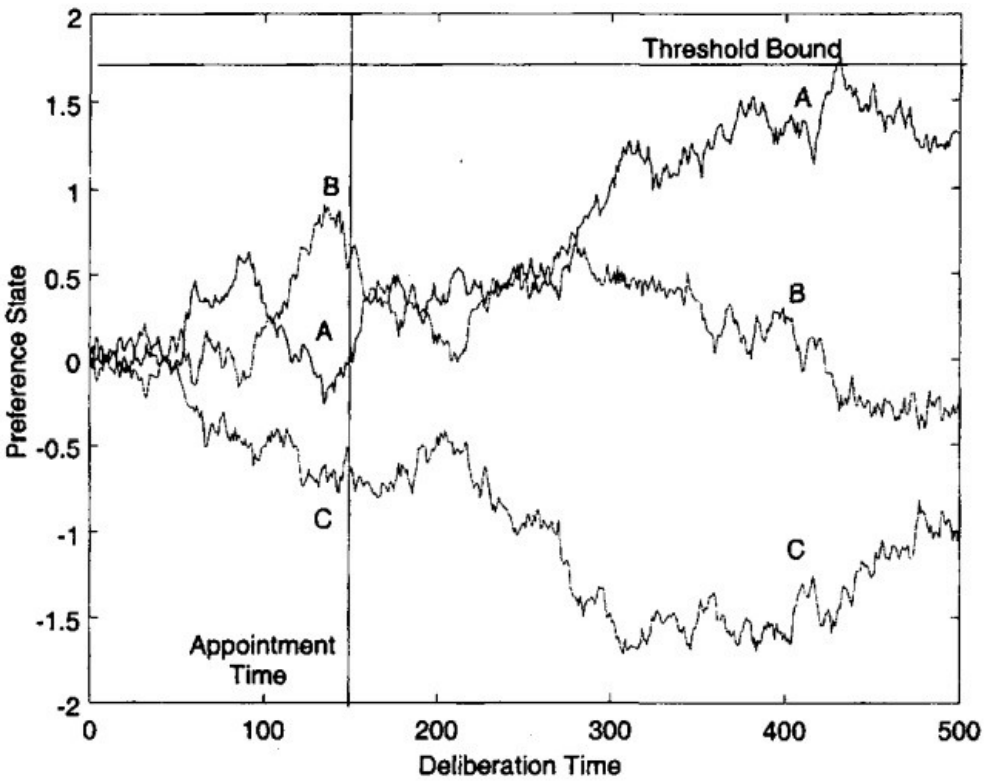


***Fig. d – Illustration of multi-alternative decision field theory with two stopping rules.***[184]

[183] *Ibid*.

[184] *Ibid*., 375.

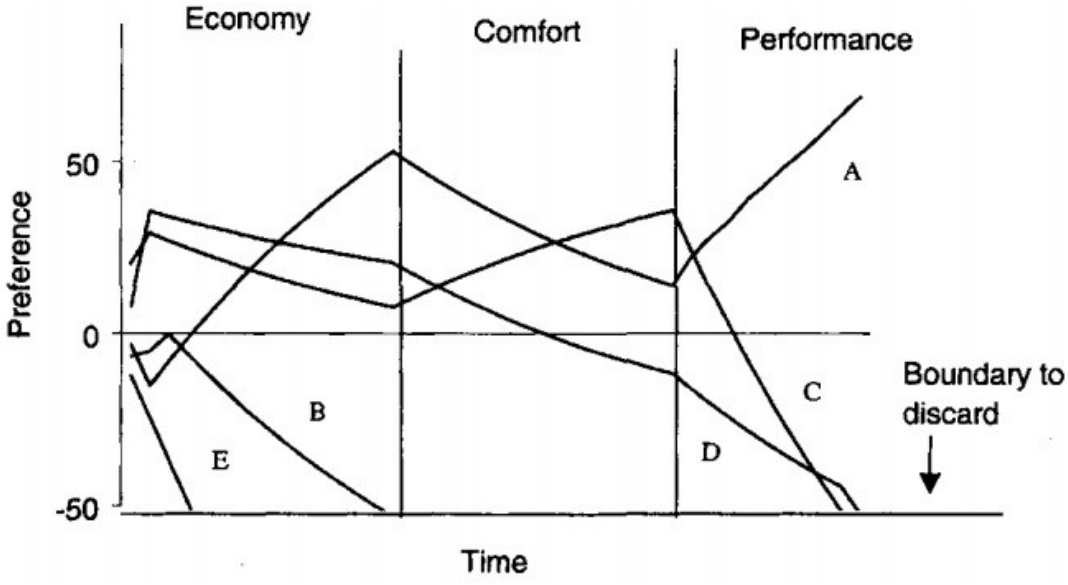


***Fig. e - Illustration of multi-alternative decision field theory with lower boundary.***[185]

As Roe, Busemeyer and Townsend explain, the 'complete version of this decision process uses both an upper acceptance boundary and a lower rejection boundary', respectively representing the requisite level of preference for any available option to be accepted as a final outcome or rejected from consideration. The authors further describe how such a double-boundary version of the theory may 'mimic strategy switching by allowing the lower reject boundary to change depending on the number of options initially presented to the decision maker.'[186] Where a given decision possesses a large number of available options, the lower boundary may be set relatively close to the "neutral" or "zero preference state" so that inferior options can be rejected swiftly. Where the number of available options are (or have been reduced to) few, the lower boundary may be positioned relatively further away from the neutral point in order to give sufficient time for each option to be properly considered and avoid its premature elimination.[187]

Models in support of the theory have largely been built upon primate experiments concerning perceptual decisions.[188] In particular, the use of primates allows for electrical

[185] *Ibid*., 385.
[186] *Ibid*.
[187] *Ibid*.
[188] Rubén Moreno-Bote, John Rinzel and Nava Rubin, 'Noise-induced alternations in an attractor network model of perceptual bistability' (2007) 98(3) *Journal of Neurophysiology* 1125; Xiao-Jing Wang, 'Probabilistic decision making by slow reverberation in cortical circuits' (2002) 36(5) *Neuron* 955.

recordings to be taken directly from neurons in the animal subject's brain, whilst it is possible to train the animals to act in particular ways in response to perceptual decisions that they have taken and concurrently record electrical activity in those neurons. For example, the primate will be trained to visually distinguish between the direction of movement of dots on a screen, indicating their decision with an eye movement (saccade). Electrical recordings thus demonstrate how 'neuronal activity is primarily correlated with the decision choice' whilst 'spike discharges build up over time, at a faster speed with stronger stimulus strength' and 'categorical choice is stored in working memory.'[189] In a similar experiment investigating motor decisions rather than perceptual, neuronal activity in the motor areas of the brain were shown to represent competing reaching actions and the selection between them. The recorded neuronal activity reliably predicted both the primate's response choices and, indeed, errors.[190]

### 2.3.2. *Integrating the Expanded Brass-Haggard Model with Multi-alternative Decision Field Theory*

As an initial step to integrating the expanded Brass-Haggard model of decision-making and multi-alternative decision field theory, Cisek first notes that, in order to 'successfully accomplish a behavioral goal such as reaching for an object, an animal must solve two related problems: to decide which object to reach and to plan the specific parameters of the movement,'[191] referring to the *what* and *how* components of a decision respectively. Whereas these have traditionally been treated as separate problems that must be solved serially – *i.e.* first decide *what* to do and then *how* to do it – Cisek offers a "Computational Model" under which populations of neurons that are tuned to specific spatial movement parameters (the *how* component) are active 'in proportion to sensory and cognitive information favouring the selection of actions with the specific parameter value' (the *how* component).[192] Consequently, this mixed representation 'can be used to solve, in parallel,

[189] Wang (2002), 964.
[190] Paul Cisek and John F. Kalaska, 'Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action' (2005) 45 (5) *Neuron* 801.
[191] Paul Cisek, 'Integrated neural processes for defining potential actions and deciding between them: A computational model' (2006) 26(38) *Journal of Neuroscience* 9761; Paul Cisek, 'Cortical mechanisms of action selection: The affordance competition hypothesis' (2007) 362(1485) *Philosophical Transactions of the Royal Society: Biological Sciences* 1585.
[192] Cisek (2006), 9761.

both the problem of specifying the spatial metrics of a potential action (an aspect of planning) and the problem of selecting between different potential actions (decision-making).' [193] This is taken to explain, in turn, why neural activity can be found simultaneously across the parietal, prefrontal and premotor cortices during decision-making, reflecting the integration of perceptual, cognitive and action planning processes.

Cisek later proceeds to consider three possible models for decision-making (***figure f***), each reflecting different ways in which a multi-alternative field theory might operate in practice. Under **(a)** a good-based model, decisions are reached by comparing the neural representations of different values for each available option and, after a particular choice has been selected, motor plans are developed to translate that decision into action. Under **(b)** an action-based model, different potential actions are represented by neural networks and a decision is reached through 'biased competition between those action representations.'[194] Finally, under **(c)** a distributed consensus model, different goals and their corresponding actions are represented across many levels, and decisions are reached through 'competition at multiple levels of representation.'[195]

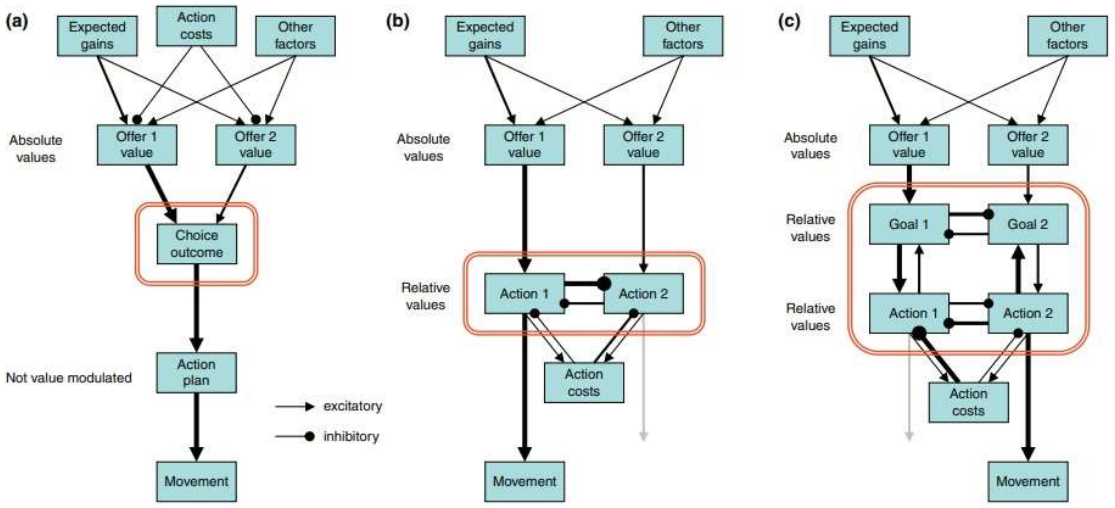


***Fig. f – Three possible schemes for deciding between actions.***[196]

[193] *Ibid*.
[194] *Ibid*.
[195] *Ibid*.
[196] Paul Cisek, 'Making decisions through a distributed consensus' (2012) 22(6) *Current Opinion in Neurobiology* 927, 928.

Beginning with the good-based model, one apparently critical flaw is that it predicts that motor planning would only begin after a substantive decision has been reached, whereas 'many studies have shown that neurons in sensorimotor regions represent multiple potential targets and actions.'[197] Indeed, Cisek argues that it is unclear how the brain might even properly compute competing action options (and, in particular, the costs associated with each option) if it does not have some representation of those competing actions, and further evidence reveals how people are 'remarkably sensitive' to the costs of different potential actions.[198] Furthermore, the good-based model does not account for why neural activity in sensorimotor regions would be modulated by variables that are more relevant to the decision itself, with neural activity relating to an action tending to be stronger where that action returns greater rewards.

Turning to the action-based model, Cisek notes that the brain evolved not to deal with purely abstract problem-solving tasks but to support natural behaviours such as hunting, foraging, and escaping dangers *etc.* He writes, 'in the natural environment, decisions between simultaneous options are usually associated with particular actions, whose metrics are specified by geometric information picked-up by the sensors.'[199] This holds true whether an animal (or person) is deciding in which direction to search for new food, or how best to escape a predator that is chasing it. Considering the latter example of escaping predation, an action-based model offers greater advantages to both predator and prey than would a good-based model; by preparing multiple action plans for either catching or escaping from another animal, these competing plans can be continuously updated by sensorimotor information until the final moment that a particular decision is taken and route chosen. Under a good-based model, such decisions would be predictably slower, as the animal would always first have to select an option (*e.g.,* go left or right)

[197] *Ibid*; citing He Cui and Richard A. Andersen, 'Posterior parietal cortex encodes autonomously selected motor plans' (2007) 56(3) *Neuron* 552; Camillo Padoa-Schioppa, 'Range-adapting representation of economic value in the orbitofrontal cortex' (2009) 29(44) *Journal of Neuroscience* 14004; Alexandre Pastor-Barnier, Elsa Tremblay and Paul Cisek, 'Dorsal premotor cortex is involved in switching motor plans' (2012) 5(5) *Frontiers in Neuroengineering* 1; Hansjörg Scherberger and Richard A. Andersen, 'Target selection signals for arm reaching in the posterior parietal cortex' (2007) 27(8) *Journal of Neuroscience* 2001.

[198] Cisek (2012), 928; citing Ignasi Cos, Nicolas Bélanger and Paul Cisek, 'The influence of predicted arm biomechanics on decision making' (2011) 105(6) *Journal of Neurophysiology* 3022.

[199] Cisek (2012), 930.

and subsequently plan the appropriate action, taking up valuable time in a predation scenario.

The critical flaw with the action-model, however, is that by its very name the model relates to decisions regarding actions, and cannot therefore explain how more abstract non-motor decisions are reached. Whereas the brain could operate two systems for making abstract and motor decisions, it is more reasonable to suppose that 'considerations of evolutionary continuity' resulted in a single system that could respond flexibly to deal with different kinds of decisions, not least considering the 'highly conservative' nature of brain evolution and the fact that mammalian decision-making mechanisms evolved over millions of years to deal almost exclusively with action-based decisions.[200] Consequently, Cisek writes how the 'challenges of a continuously changing environment demanded the evolution of a functional architecture in which the mechanisms specifying possible actions and those which evaluate how to select between them can operate in parallel.'[201] This can be read as describing the expanded Brass-Haggard model, whereunder the various components of a decision are processed in parallel by at least partially distinct networks in the brain.

Referring to the distributed consensus model depicted at **(c)** in ***figure f***, competition occurs between neuronal networks representing goals and actions on multiple levels, thereby accounting for both abstract and motor decisions. As Cisek describes, activity on the lower level of the diagram reflects competing actions, whilst activity on the higher level reflects competing choices; however, integrating the expanded Brass-Haggard model, it is posited that further levels could simultaneously further represent the *when*, *whether* and *why* components of any given decision. Diagram **(c)** at ***figure f*** also shows various excitatory and inhibitory linkages between the different levels, but these need not necessarily be one-to-one and many goals may lead to a single action or *vice versa*. Cisek explains, 'because the levels are reciprocally connected, they share the biases that may arrive from a variety of sources, and gradually arrive at a decision through a "distributed

[200] *Ibid*.
[201] *Ibid*.

consensus".'[202] Different decision and action choices may be biased by a range of factors impacting upon their relative overall values, such as action costs, remembered and predicted values of each option, sensorimotor contingencies *etc.* Whereas such biases may not be in agreement with regards to a given decision, 'positive feedback between the layers will eventually force a choice to emerge.'[203]

This, therefore, describes how the expanded Brass-Haggard model can be integrated with multi-alternative decision field theory. The various *what, how, when, whether* and *why* components of a decision are processed in parallel across at least partially distinct neural networks in various regions of the brain. These each draw from relevant biases, such as geometric and sensorimotor information biasing competing *how* components; remembered and predicted values and costs biasing competing *what* components, *etc*; as well as different components feeding back into one another, such as the predicted values for different action plans (*how* component) feeding back into the respective values of different goals (*what* component). These various biases may not necessarily all point towards to same choice; yet, as the various networks are only partially distinct, linkages and positive feedback between each network eventually allows for a decision to be made according to consensus for a single option being reached across the multiple layers.

*

From the integrated model of decision-making described in this section of the thesis, one potential role for consciousness in the decision-making processes may be hypothesised. Specifically, the integrated model describes at least five different but interrelated processes – the *what, how, when, whether* and *why* components of any single decision – which are processed in parallel by the brain until a point of distributed consensus is reached. However, our interaction with the world and, indeed, our conscious experience exists largely in serial; we generally have one stream of conscious, think one thought at a time, and can meaningfully engage in one action at a time. Although people may sometimes be able to engage in multiple thought processes or activities, these are often

[202] *Ibid.,* 932
[203] *Ibid.,* 933.

trivial such as conversing about one topic whilst carrying out some manual activity with the hands, or writing an e-mail whilst watching television in the background. However, significant and meaningful tasks requiring concentration can typically only be properly and effectively carried out with concentration and focus on one thing at a time.

It might be hypothesised, therefore, that one of the roles played by consciousness in decision-making is to act as an interface, translating the multiple parallel processes necessary for any given decision into a single, serial experience through which that decision can be rendered into action. As the experience of interaction with and, in particular, *reaction* to the outside world occurs serially with conscious thought and broad motor responses occurring one at a time, consciousness itself may be a necessary component to allow such underlying multiple parallel processes to produce the singular serial experience necessary for real-world interaction. It might be further posited that without the translation of parallel processes into serial experience through consciousness, an animal attempting to act (or react) in the world could be effectively paralysed by the inability to translate those multiple parallel processes into a singular action that can be performed by the body, or indeed by an inability to hierarchically arrange competing actions into the serial order that they need to be performed by the body.

## *2.4. From the Science of Decision-Making to Legal Responsibility*

As outlined briefly in the introduction to this thesis, current descriptions of legal responsibility require that an individual commits a prohibited act or omission (*actus reus*) with the requisite subjective state of mind (*mens rea*), and in the absence of any exculpatory or justificatory factors such as coercion and duress, lack of consciousness (automatism) and self-defence *etc.*[204] The capacity to 'grasp and be guided by good reason' is presumed to exist for all adults unless some relevant defence is raised which may negate this presumption, whilst it is further presumed that this capacity in turn

[204] Jeremy Horder, *Ashworth's Principles of Criminal Law* (9th ed. Oxford University Press 2019), Chs. 5 – 7.

enables people to decide and act volitionally, *i.e.,* to exert conscious control over their decisions and actions.[205]

Every criminally-relevant action (*actus reus*) begins with a decision to act. This is not least reflected in the fact that the law does not generally punish accidents, reflexes, and other actions performed without conscious control, which is itself reflected in a number of defences. For example, the defence of diminished responsibility requires that the defendant was suffering from an 'abnormality of mental functioning, arising from a recognized medical condition, which… substantially impaired his or her ability to understand the nature of his or her conduct, form a rational judgment, and exercise self-control.'[206] The defence of diminished responsibility is available where the defendant's mental abnormality is not so severe as to attract defences of insanity or automatism. It does not denote that the defendant is entirely unable to act voluntarily, but rather that their perceptions or character are 'so distorted that he is unable critically to evaluate his conduct.'[207]

For an insanity defence, the defendant must demonstrate that they suffered at the time of the offence from a 'defect of reason caused by a disease of the mind which meant that either: (1) he or she did not know the nature or quality of his or her actions; or (2) he or she did not know that what he or she was doing was wrong.'[208] Here, "disease of the mind" attracts a normal (as opposed to specifically medical) interpretation, and the disease itself need not be a psychiatric disorder;[209] for example, brain malfunctioning caused by diabetes will be considered a disease of the mind. Crucially, the second component of the insanity defence again reflects the fact that the defendant no longer possessed the requisite understanding or control over their actions. This component may be satisfied where the

[205] Stephen J. Morse, 'Moral and legal responsibility and the new neuroscience' in Iles J. (ed.), *Neuroethics: Defining the Issues in Theory, Practice, and Policy* (Oxford University Press 2006), 37 – 38; Stephen J. Morse, 'The non-problem of free will in forensic psychiatry and psychology' (2007) 25(2) *Behavioral Sciences & the Law* 203.
[206] Jonathan Herring, *Criminal Law: Text, Cases, and Materials* (9th ed. Oxford University Press 2020), 254 – 255; citing Homicide Act 1957, s. 2(1) (as amended by the Coroners and Justice Act 2009).
[207] *Ibid*., 319; citing Ronnie D. Mackay, 'Diminished responsibility and mentally disordered killers' in Ashworth A. and Mitchell B. (eds.), *Rethinking English Homicide Law* (Oxford University Press 2000).
[208] *Ibid*., 666; *R v M'Naughten* (1843) 8 ER 718; *R v Sullivan* [1984] 1 AC 156.
[209] *R v Kemp* [1957] 1 QB 399, 406.

defendant had no conscious awareness of what was happening (such as during a seizure), where they were 'deluded as to the material circumstances' of their actions, or where they were unaware of the consequences of their actions.[210]

At the most extreme end of the scale, the defence of automatism asserts that the defendant suffered from a '*complete* loss of voluntary control' due to some *external* cause,[211] whereas automatism arising from some *internal* cause is more properly pleaded under the defence of insanity.[212] As may readily be discerned, each of these defences is concerned with the degree to which some relevant factor – *i.e.* medical condition or abnormality of the mind – changes, diminishes or entirely abrogates conscious, volitional control over actions. As is stated in the seminal case of *R v M'Naughten*, all adult defendants are first presumed to be sane and, therefore, it is equally presumed that any given adult defendant possesses the requisite capacities for rational thought and volitional control over their actions, unless the contrary is demonstrated such as through raising one of the aforementioned defences.

In order to attract legal responsibility, the decision which precedes an *actus reus* must fall within one of a number of particular states of mind – *mens rea* – which are reasoned to denote moral fault on the guilty individual.[213] For example, the offence of battery requires that the defendant *intended* to touch or hit another individual as opposed to accidentally bumping into them on in a crowded environment. Similarly, the offence of theft requires that the defendant *dishonestly* took property belonging to another, as opposed to mistakenly picking up somebody else's bag which they believed to be their own. Thus, *mens rea,* as it is traditionally understood, describes the subjective mental states which

[210] Herring (2020), 698.
[211] *Ibid*., 690; citing *Attorney-General's Reference (No. 2 of 1992)* [1994] QB 91.
[212] *R v Sullivan* [1984].
[213] David Ormerod and Karl Laird, *Smith, Hogan, and Ormerod's Text, Cases, and Materials on Criminal Law* (13th ed. Oxford University Press 2020), 96. In reality, not all *mens rea* can truly be described as reflecting a particular state of mind; for example, crimes committed by negligence require that there was a breach of some legal duty, which is not concerned with the state of mind of the accused. Nonetheless, *mens rea* is generally discussed as reflected the "mental" element of criminal offences, in contrast to the *actus reus* "action" element.

denote moral fault and blameworthiness upon the guilty party for committing a related *actus reus*.

Where every criminally-relevant action (*actus reus*) begins with a decision to act formed under a particular state of mind (*mens rea*), it follows from the present chapter of this thesis that the content of any such legally-relevant decision is comprised of the five components previously discussed – *what, how, when, whether* and *why* do to a particular criminal act. That is to say, any criminal act must itself comprise of a decision as to *what* (criminal) thing to do, *how* to carry out that act, *when* to perform it, *whether* or not to go through with that act, and the reasons *why* to commit that act. The law presumes that people are able to consider and evaluate reasons in deciding to act and, further, are able to consciously control both the outcome of their decisions and their subsequent bodily actions in carrying out those decisions.

The following five chapters of this thesis explore each of these components in turn, presenting evidence from neuroscience and psychology that pertains to the relationship between scientific models of decision-making, the concept of *mens rea*, and the accompanying presumptions of the capacity for rational thought and conscious control over decisions and actions. A central consideration throughout the following chapters is the veracity of these legal presumptions – *i.e.,* can people control the outcome of their decisions; do they reach decisions rationally by recognising and responding to good and bad reasons for different options; and what is the involvement of active, conscious thought in decisions to act. Further, the following chapters consider the implications of scientific research concerning each of the five decision-making components for the law's underlying philosophical assumption that people make decisions with free will.

# 3. The *What* Component, Priming and Predicting Choices

> 'Are decisions voluntary? Or are they things that happen to us? From some fleeting vantage points they seem to be the preeminently voluntary moves in our lives, the instants at which we exercise our agency to the fullest. But those same decisions can also be seen to be strangely out of our control. We have to wait to see how we are going to decide something, and when we do decide, our decision bubbles up to consciousness from we know not where. We do not witness it being *made*; we witness its *arrival*.'
>
> - Daniel C. Dennett, 2015.[1]

The previous chapter of this thesis discussed how even the simplest decisions can be broken down into at least five components – the *what, how, when, whether,* and *why* to decide any particular thing and take resultant action. The first to consider in greater detail is the *what* component; indeed, the decision of *what* to do will often be the essence of many criminal offences – the decisions to *steal*, to *attack*, to *kill*, to *defraud*, *etc.* are all decisions about *what* to do in a particular situation. The law punishes such actions when they are committed intentionally – a form of *mens rea* – on the premise that individuals freely and consciously choose to pursue a given criminal action when they could otherwise refrain from doing so. Whilst this latter aspect of refraining from a particular action relates to the *whether* component of a decision – *i.e., whether* or not to implement a decision into action – the former aspect of deciding to pursue a given course of conduct in the first place concerns the *what* component, which is the subject of this chapter.

[1] Daniel C. Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting* (2nd ed. Massachusetts Institute of Technology Press 2015), 85.

## *3.1. Priming and Automaticity*

As pioneers in the field Bargh and Chartrand explain, research techniques focusing on priming and automaticity 'share a concern with the ways that internal mental states mediate, in a passive and hidden manner, the effects of the social environment on psychological and behavioral responses.'[2] Automaticity techniques allow for measuring mental procedures or representations that are theoretically assumed to correspond with individual phenomenological differences. Meanwhile priming studies investigate the impact of situational context and environmental features on the ways in which individuals think, feel and behave.[3] Described most simply, the phenomenon of priming may be understood as occurring when the 'processing of a stimulus is changed following presentation of another stimulus,'[4] or where there is a 'change in the response to a stimulus, or in the ability to identify a stimulus, following prior exposure to that stimulus.'[5] Eiser describes more fully:

> 'A *prime (noun)* is any piece of information, word or stimulus, typically with symbolic meaning, presented to an individual that can set in train, *i.e., prime (verb)*, a string of associations so that other concepts, thoughts or memories are more likely to come to mind and/or be acted upon.'[6]

Lashley first used the term priming in 1951, discussed within the context of behavioural priming.[7] In particular, Lashley was considering the question of how serial sequences of behaviour, such as speech, flow so swiftly and with apparently little effort required. Rejecting a then-dominant behaviouralist account of behaviour as a reflex to stimuli, Lashley argued that there exists a mediating state between the formation of an intention (such as to perform an action or speak a sentence) and the execution of that intention

---

[2] John A. Bargh and Tanya L. Chartrand, 'The mind in the middle: A practical guide to priming and automaticity research' in Reis H. T. and Judd C. M. (eds.), *Handbook of Research Methods in Social and Personality Psychology* (2nd ed. Cambridge University Press 2014), 312.
[3] *Ibid*.
[4] Graham Richards, *Psychology: The Key Concepts* (Routledge 2009), 183 – 184.
[5] Michael S. Gazzaniga, Richard B. Ivry and George R. Mangun, *Cognitive Neuroscience: The Biology of the Mind* (5th ed. W. W. Norton & Co. 2019), 392.
[6] J. Richard Eiser, 'A History of Social Judgment Research' in Kruglanski A. W. and Stroebe W. (eds.), *Handbook of the History of Social Psychology* (Psychology Press 2012), 230.
[7] Karl S. Lashley, 'The problem of serial order in behavior' in Jeffress L. A. (ed.), *Cerebral Mechanisms in Behavior* (Wiley 1951).

through overt behaviour, and that this mediating state assembled the relevant behavioural actions to execute that intention into the appropriate serial sequence. Lashley referred to this as "priming" a response: 'prior to the internal or overt execution of the sentence, an aggregate of word units is partially activated or readied.'[8]

Bargh describes the 'serendipitous' discovery of 'carryover priming effects' which, unlike the earlier Meyer and Schvaneveldt archetype,[9] lasted several minutes as opposed to a few seconds.[10] In 1958, Storms gave experimental subjects a list of words to memorise, followed later by a free-association task where the subjects would freely report words associated with stimulus words.[11] Storms reported that those words that had been presented in the memory task were more likely than other words to be given by subjects in the free-association task. In 1960, Segal and Cofer replicated Storms' experiment and referred to this effect as "priming", whereby the use of a given concept in one task increased the probability of its further use in subsequent and unrelated tasks performed shortly thereafter.[12] Priming was thereafter adopted as an experimental technique, initially to demonstrate how information could be stored in memory without an individual explicitly being able to recall that information,[13] leading to the contemporary distinction between implicit and explicit memory.[14]

Again, it is reiterated that these priming effects as originally described were "carryover effects" which lasted for several minutes and impacted upon subsequent and unrelated

[8] *Ibid*., 119; John A. Bargh, 'The historical origins of priming as the preparation of behavioral responses: Unconscious carryover and contextual influences of real-world importance' (2014) 32(Supp) *Social Cognition* 209, 211 – 212.
[9] David E. Meyer and Roger W. Schvaneveldt, 'Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations' (1971) 90(2) *Journal of Experimental Psychology* 227.
[10] Bargh (2014), 212 – 213.
[11] Lowell H. Storms, 'Apparent backward association: A situational effect' (1958) 55(4) *Experimental Psychology* 390.
[12] Sydney J. Segal and C. N. Cofer, 'The effect of recency and recall on word association' (1960) 15 *American Psychologist* 451.
[13] Stanley Grand and Sydney J. Segal, 'Recovery in the absence of recall: An investigation of color-word interference' (1966) 72(1) *Journal of Experimental Psychology* 138; Asher Koriat and Nili Feuerstein, 'The recovery of incidentally acquired information' (1976) 40(6) *Acta Psychologica* 463; Sydney J. Segal, 'The priming of association test responses: Generalizing the phenomenon' (1967) 6(2) *Journal of Verbal Learning and Verbal Behavior* 216.
[14] Daniel L. Schacter, 'Implicit memory: History and current status' (1987) 13(3) *Journal of Experimental Psychology: Learning, Memory and Cognition* 501.

tasks, in contrast to the short-lived lexical priming effects described by Meyer and Schvaneveldt which became the archetype for social and behavioural priming sceptics. In 1977, Higgins, Rholes and Jones presented a ground-breaking study demonstrating carryover priming effects within the distinctly social arena of forming impressions of other people.[15] Subjects were first primed with certain personality traits by being exposed to synonyms for those traits in a memory task. Second, subjects were presented with descriptions of target people whose behaviour was described in ambiguous ways; for example, accounts of a person sailing alone across the ocean or preferring to study alone as opposed to with others could be interpreted as being "adventurous" and "independent" or, conversely, "reckless" and "aloof". Those subjects who had been primed with positive traits subsequently formed a more positive impression of the target character, whilst those subjects primed with negative traits formed a more negative impression. Importantly, subjects did not report any subjective awareness of being influenced by the first memory task in their impressions formed on the second social task. As Bargh comments:

> 'The… study revealed for the first time how an individual's recent experience could affect, in a passive and unintended manner, his or her perceptual interpretation of another person's behavior. In their study, all participants read about the same target person doing the same things, yet they came away from their reading with markedly different impressions of that person, differences that were only accountable by reference to the experimentally manipulated differences in their recent use of different trait concepts.'[16]

The following decades witnessed an explosion of priming research across a whole range of areas that are hereby summarised. *Perceptual priming* describes where a priming stimulus influences upon the perception (*i.e.,* the detection or identification) of a

[15] E. Tory Higgins, William S. Rholes and Carl R. Jones, 'Category accessibility and impression formation' (1977) 13(2) *Journal of Experimental Social Psychology* 141; replicated by Thomas K. Srull and Robert S. Wyer, 'The role of category accessibility in the interpretation of information about persons: Some determinants and implications' (1979) 37(10) *Journal of Personality and Social Psychology* 1660; John A. Bargh, Ronald N. Bond, Wendy J. Lombardi and Mary E. Tota, 'The additive nature of chronic and temporary sources of construct accessibility' (1986) 50(5) *Journal of Personality and Social Psychology* 869.
[16] Bargh (2014), 213.

subsequent stimulus.[17] The paradigmatic experimental example consists of the word-fragment completion task in which subjects are first exposed to priming words without realising their significance. Second, subjects perform a task in which partial word-fragments are provided and they must fill in letters to complete the fragment and make a word. Subjects are subsequently more likely to complete the word-stems with words to which they were previously exposed, revealing a priming effect from their prior perception of those words.[18]

Whereas perceptual priming is eponymously concerned with the perceptual properties of a prime such as its auditory or visual appearance, *conceptual* or *semantic priming* is concerned with meaning behind a prime. For example, where the presentation of a "table" may prime the perception of a table in a subsequent word task, it may also prime a "chair" because the concepts of table and chair are semantically related.[19] In Meyer and Schvaneveldt's famous 1971 experiment,[20] subjects were provided with word pairs such as "table-grass", half of which were semantically related such as "nurse-doctor". Subjects were subsequently faster in responding to semantically related word-pairs in a subsequent task. The aforementioned paradigm by Higgins, Rholes and Jones also uses conceptual priming; after subjects were first primed with semantically related characteristics before reading the ambiguous description of a character, they were subsequently more likely to form an impression of that character similar to the semantic primes.[21]

---

[17] See Daniel L. Schacter, 'Priming and multiple memory systems: Perceptual mechanisms of implicit memory' in Schacter D. L. and Tulving E. (eds.), *Memory Systems* (Massachusetts Institute of Technology Press 1994); Endel Tulving and Daniel L. Schacter, 'Priming and human memory systems' (1990) 247(4940) *Science* 301.

[18] Endel Tulving, Daniel L. Schacter and Heather A. Stark, 'Priming effects in word-fragment completion are independent of recognition memory' (1982) 8(4) *Journal of Experimental Psychology: Learning, Memory and Cognition* 336; Peter Graf, George Mandler and Patricia E. Haden, 'Simulating amnesiac symptoms in normal subjects' (1982) 218(4578) *Science* 1243.

[19] See Timothy P. McNamara, *Semantic Priming: Perspectives from Memory and Word Recognition* (Psychology Press 2005), 3 – 9.

[20] Meyer and Schvaneveldt (1971).

[21] See further James H. Neely, 'Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention' (1977) 106(3) *Journal of Experimental Psychology* 226; James H. Neely, 'S Semantic priming effects in visual word recognition: A selective review of current findings and theories' in Besner D. and Humphreys G. W. (eds.), *Basic Processes in Reading: Visual Word Recognition* (Lawrence Erlbaum Associates 1991).

Behavioural priming goes beyond perceptual and semantic priming with regards to how primes do not merely activate perceptions and concepts in the mind, but can further influence subsequent behaviours and motor actions. The quintessential example of behavioural priming within academic literature is arguably a paradigm by Bargh, Chen and Burrows.[22] Subjects were first given a scrambled-sentence task in which they had to rearrange scrambled sets of five words in order to create grammatically correct four-word sentences. This task was used to prime subjects with words semantically related to the concept of "elderly", including such priming words as "Florida", "bingo", "retired", "wrinkle", "traditional" and "ancient". After completing the task and being given a fake debriefing, the subjects were timed by a confederate with a hidden stopwatch as they walked from the experimental room to an elevator down the hall. The results showed that subjects who were primed with the elderly concept walked significantly slower from the experimental room to the elevator than those who had not been so primed.

Finally, social priming is a less precise term which refers to priming demonstrated within a social context or task. The aforementioned example of semantic priming and forming an impression of an ambiguously described character also falls within the category of social priming, as the task of forming a character impression is an inherently social activity. Similarly, the aforementioned study by Bargh, Chen and Burrows included two further experiments in social priming. Subjects who were primed with either of the concepts of rudeness or politeness waited for correspondingly shorter or longer times before interrupting a conversation between the experimenters. Similarly, Caucasian subjects primed with images of African-American faces displayed greater aggression towards an annoying request by the experimenter than those who had been primed with Caucasian faces.

It is important to note that priming effects are not proposed to be overwhelming or inescapable on human decision-making and behaviour. Rather, primes operate subtly and are highly contextual; that which primes one individual may not necessarily prime another

[22] John A. Bargh, Mark Chen and Lara Burrows, 'Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action' (1996) 71(2) *Journal of Personality and Social Psychology* 230.

in the same circumstances. Indeed, this is entirely logical; if priming exerted an overbearing effect on behaviour, people would become slaves to every potential influence or advertisement that crossed their path, which is patently not the case. However, like commercial advertisements, primes can nonetheless exert a critical "nudge" in people's decision-making with undoubtedly significant impacts upon their subsequent behaviour. Perhaps most importantly for the purposes of the present thesis, the phenomenon of priming reveals how decision-making and behavioural process can and do operate both automatically and unconsciously in the mind.

### *3.1.1. Priming Responses Outside of Awareness*

This first section considers studies where subjects' responses to certain scenarios or circumstances have been primed, as distinct from the following section which considers a subtly different phenomenon of priming actual goals, intentions or motivations. Priming responses is potentially most relevant with regards to offences that arise and are committed opportunistically; for example, the thief who takes an unattended phone from a café table, as opposed to one who plans and executes a violent robbery of the till. The argument follows that if people may be primed to respond in certain ways – *e.g.,* to opportunistically take property that has been left unattended, or to respond aggressively to confrontation – and if such priming can occur and operate outside of conscious awareness and control, this may have important implications for notions of responsibility. An important precursory note is that, unless stated otherwise, references in this section to priming "unconsciously" or "outside of awareness" do not mean that subjects are unaware of the stimulus that is priming them *i.e.,* that priming is subliminal. Rather, the lack of awareness refers to subjects being unaware of the influence that a priming stimulus may have on their responses. Thus, when priming subjects with the faces of people from different races, for example, subjects are entirely aware that they observing a face, but they will typically be unaware of the priming effect that stimulus may have on their subsequent decisions or behaviour.

An apt starting point is research concerning the priming of hostile or aggressive responses, as it is most obvious how somebody primed in such a way might more readily descend

into aggressive or violent criminal behaviour. The previous section of this thesis introduced the seminal studies by Bargh, Chen and Burrows, which included one experiment where Caucasian subjects primed with African-American faces displayed greater aggression when later asked by the experimenter to fulfil a frustrating request.[23] Relatedly, Brown, Croizet, Bohner, Fournet and Payne demonstrated that priming the African-American stereotype resulted in decreased cooperation amongst individuals who were already highly prejudiced.[24] (Similar displays of priming based on racial stereotypes include a study by Wheeler, Jarvis and Petty in which subjects' performance on maths tests deteriorated after priming with African-American stereotypes,[25] whilst Shih, Ambady, Richeson, Fujita and Gray demonstrated an improvement on maths test performance when subjects were primed with Asian-American stereotypes).[26] One critical moderator with regards to priming racial stereotypes is naturally the extent and degree to which subjects hold such racial stereotypes or prejudice in the first place; those subjects not holding such stereotypes were predictably not primed with them.

The "weapons priming effect" is one of the most well-documented examples of priming aggressive responses. In 1967, Berkowitz and LePage[27] conducted an experiment where male subjects first received between 1 and 7 mild electric shots, supposedly administered by one of their peers. Subjects then had the opportunity to purportedly return electric shocks to those peers; for test subjects there were guns on the table next to the shock key which the subject was told belonged to their target, whilst for control subjects there were badminton racquets on the table instead or nothing at all. Subjects who had received more shocks administered more electric shocks in return when there were weapons on the table,

[23] Whereas the experiment within the same study concerning walking speed after priming the stereotype of "elderly" has failed to replicate, there is no similar reported failure to replicate the experiment priming aggression with racial stereotypes.

[24] Rupert Brown, Jean-Claude Croizet, Gerd Bohner, Marion Fournet and Andrew Payne, 'Automatic category activation and social behaviour: The moderating role of prejudiced beliefs' (2003) 21(3) *Social Cognition* 167.

[25] S. Christian Wheeler, W. Blair G. Jarvis and Richard E. Petty, 'Think unto others: The self-destructive impact of negative racial stereotypes' (2001) 37(2) *Journal of Experimental Social Psychology* 173.

[26] Margaret Shih, Nalini Ambady, Jennifer A. Richeson, Kentaro Fujita and Heather M. Gray, 'Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation' (2002) 83(3) *Journal of Personality and Social Psychology* 638.

[27] Leonard Berkowitz and Anthony LePage, 'Weapons as aggression-eliciting stimuli' (1967) 7(2) *Journal of Personality and Psychology* 202.

and yet more shocks still when they were told that the weapons belonged to the peer who had shocked them.[28]

The weapons effect – referring to the effect of the mere presence of weapons increasing aggressive responses to other unrelated stimuli – was replicated numerous times in the following decades. Anderson, Benjamin and Bartholow presented a further experiment in 1998 strengthening the link between the weapons effect and the phenomenon of priming generally. [29] Subjects primed with weapon or non-weapon stimuli subsequently completed a word-pronunciation task where they had to repeat a word aloud as soon as it was presented. The results found a significant increase in the speed of naming aggressive words after exposure to weapons-related words and images, confirming the hypothesis that weapons primed other semantic concepts related to aggression and / or violence.

The weapons priming effect has been replicated robustly and enjoys further support from multiple meta-analytic reviews.[30] Interpreted at its high point, prominent researchers in the field Benjamin and Bushman submit that the 'link between weapons and aggression is very strong in semantic memory, and that merely seeing a weapon can make people more aggressive.'[31] This research has also revealed a number of moderators of the effect; for example, the effect has been demonstrated in adolescents viewing pictures of weapons, but only in boys who already displayed assessed high-aggressiveness, suggesting that gender and aggressive personality traits are relevant moderators. [32] However, a particularly illuminating study by Bartholow, Anderson, Carnagey and Benjamin illustrates the complex interplay between primes and their moderating factors, comparing

[28] *Ibid*., 205.
[29] Craig A. Anderson, Arlin J. Benjamin and Bruce D. Bartholow, 'Does the gun pull the trigger? Automatic priming effects of weapon pictures and weapon names' (1998) 9(4) *Psychological Science* 308.
[30] Michael Carlson, Amy Marcus-Newhall and Norman Miller, 'Effects of situational aggression cues: A quantitative review' (1990) 58(4) *Journal of Personality and Social Psychology* 622; Arlin J. Benjamin, Sven Kepes and Brad J. Bushman, 'Effects of weapons on aggressive thoughts, angry feelings, hostile appraisals, and aggressive behavior: A meta-analytic review of the weapons effect literature' (2018) 22(4) *Personality and Social Psychology Review* 347.
[31] Arlin J. Benjamin and Brad J. Bushman, 'The weapons priming effect' (2016) 12 *Current Opinion in Psychology* 45, 45.
[32] Qian Zhang, JingJin Tian, Jian Cao, Da-Jun Zhang and Philip Rodkin, 'Exposure to weapon pictures and subsequent aggression during adolescence' (2016) 90 *Personality and Individual Differences* 113.

subjects who were hunters (and therefore possessed more detailed and specific experience and knowledge surrounding guns) with non-hunters.[33]

A first experiment revealed the more detailed knowledge of hunters as well as the interaction between hunting experience and types of guns (*i.e.,* hunting rifles and assault weapons) which predicted affective and cognitive reactions to guns. A second experiment demonstrated that the weapons effect on aggression was more likely to be elicited in hunters than non-hunters by stimuli depicting assault weapons, whereas the same effect was more likely in non-hunters than hunters exposed to images of hunting rifles. A third experiment revealed how the priming effect was further moderated by differences in affective and cognitive responses to the weapon cues. Thus, the study demonstrates a complex relationship between the types of weapons being used as cues, the experience of individual subjects using or not using guns, and the different cognitive and affective responses of subjects to weapons.

One notably more contentious area of research concerning the priming of aggression investigates whether violence in the media – in particular film and television, videogames, and advertisements – can also prime subsequent aggressive behaviour in the viewers of such media.[34] An initial illuminating study by Josephson in 1987[35] exposed boys aged 7 to 9 years either to violent or non-violent television segments pre-tested to be equally exciting and arousing; the violent segments included SWAT team members using walkie-talkies, which served as a violence-related cue. This was followed by a frustration procedure consisting of a short cartoon which became increasingly interrupted by static, designed to induce frustration in the subjects. The subjects were then taken to play a game of indoor hockey, immediately before which each conducted a short pre-match interview;

[33] Bruce D. Bartholow, Craig A. Anderson, Nicholas L. Carnagey and Arlin J. Benjamin, 'Interactive effects of life experience and situational cues on aggression: The weapons priming effect in hunters and non-hunters' (2005) 41(1) *Journal of Experimental Social Psychology* 48.

[34] See generally Brad J. Bushman, L. Rowell Huesmann and Jodi L. Whitaker, 'Violent media effects' in Nabi R. L. and Oliver M. B. (eds.), *The SAGE Handbook of Media Processes and Effects* (SAGE Publications 2009), 361 – 365 & 367 – 369; L. Rowell Huesmann, Eric F. Dubow and Grace Yang, 'Why it is hard to believe that media violence causes aggression' in Dill K. E. (ed.), *The Oxford Handbook of Media Psychology* (Oxford University Press 2013) 159 – 162.

[35] Wendy L. Josephson, 'Television violence and children's aggression: Testing the priming, social script, and disinhibition predictions' (1987) 53(5) *Journal of Personality and Social Psychology* 882.

the subjects were split such that half of those previously exposed to violent television segments could also see a walkie-talkie (the violence cue) during the interview. Observers who were blind to the experimental conditions then observed the children as they played three rounds of hockey, spotting for signs of aggression from any players such as shoving others, being violent with the hockey sticks, or issuing verbal abuse.

First amongst the findings, viewing violent television content did increase the aggressive behaviour of subjects in the subsequent hockey match, but only for those boys who had received higher scores of characteristic aggressiveness during pre-testing before the experiment. Second, those characteristically more aggressive subjects displayed yet more aggressive behaviour if they had been exposed to the violent television segments *and* the walkie-talkie cue immediately prior to playing hockey. Third, the additional aggressive behaviour was exhibited immediately during the first round (3 minutes) of hockey, but then dissipated thereafter.[36] Josephson explains the reported effects in the context of priming and priming moderators. Thus, the exposure to violent television segments (and the subsequent violence cue) activates concepts and memories related to aggression, whilst the 'activation of aggressive thoughts, feelings, and action tendencies would lead to actual aggressive behavior among those boys who have an established history of interpersonal aggression.'[37]

Two experiments by Anderson in 1997 revealed similar findings with subjects who were university undergraduates.[38] Subjects were randomly assigned to watch either violent fight scenes from movies or equally interesting non-violent scenes, following which they completed a questionnaire and a reaction time task where they had to repeat words as quickly as possible which could be related to concepts of aggression, anxiety, escape or control. The first experiment revealed that all subjects viewing the violent scenes later self-reported higher feelings of a state of hostility during the follow-up questionnaire.[39] In the second experiment, the moderating factor of individual trait hostility was included

[36] *Ibid*., 888.
[37] *Ibid*.
[38] Craig A. Anderson, 'Effects of violent movies and trait hostility on hostile feelings and aggressive thoughts' (1997) 23(3) *Aggressive Behavior* 161.
[39] *Ibid*., 168 – 169.

to reveal that those subjects with assessed low-trait hostility had significantly faster reaction times to repeating words related to aggression after priming with violent movie scenes. Thus, together the experiments displayed how violent film media can prime both aggressive feelings and thoughts, with initial trait hostility again being revealed as an important moderator of the priming effect.

The evidence for an aggression priming effect from violent film and television media is robust,[40] and receives the additional support of a number of meta-analytical studies.[41] Of perhaps even greater contention, however – and not least within the popular media – is the possibility of priming aggression from playing videogames. The available research on priming from videogames is comparatively smaller than that relating to film and television owing to the relative ages of the different forms of media. That notwithstanding, a mounting body of research does similarly suggest aggression priming effects from playing violent videogames. For example, three experimental studies, one correlational study and a meta-analysis published together by Anderson, Carnagey, Flanagan, Benjamin, Eubanks and Valentine each provide compelling support for aggression priming effects from violent videogames.[42] The experiments demonstrated that playing violent videogames did indeed increase the general accessibility of aggressive thoughts and behaviours as measured in a competitive reaction time task, once again indicating initial trait hostility as a key moderating factor.[43] Meanwhile, the meta-analysis confirmed significant effects of playing violent videogames on subsequent 'aggressive

[40] For example, see further Brad J. Bushman, 'Moderating role of trait aggressiveness in the effects of violent media on aggression' (1995) 69(5) *Journal of Personality and Social Psychology* 950; Brad J. Bushman, 'Priming effects of media violence on the accessibility of aggressive constructs in memory' (1998) 24(5) *Personality and Social Psychology Bulletin* 537; Sarah M. Coyne, Jennifer Ruh Linder, David A. Nelson and Douglas A. Gentile, '"Frenemies, Fraitors, and Mean-em-aitors": Priming effects of viewing physical and relational aggression in the media on women' (2012) 38(2) *Aggressive Behavior* 141; Zhang Qian, Dajun Zhang and Lixin Wang, 'Is aggressive trait responsible for violence? Priming effects of aggressive words and violent movies' (2013) 4(2) *Psychology* 96.

[41] Haejung Paik and George Comstock, 'The effects of television violence on antisocial behavior: A meta-analysis' (1994) 21(4) *Communication Research* 516; Brad J. Bushman and Craig A. Anderson, 'Media violence and the American public: Scientific fact versus media misinformation' (2001) 56(6/7) *American Psychologist* 477; David R. Roskos-Ewoldsen, Mark R. Klinger and Beverly Roskos-Ewoldsen, 'Media priming: A meta-analysis' in Preiss R. W., Gayle B. M., Burrell N., Allen M. and Bryant J. (eds.), *Mass Media Effects Research: Advances Through Meta-Analysis* (Routledge 2007).

[42] Craig A. Anderson, Nicholas L. Carnagey, Mindy Flanagan, Arlin J. Benjamin, Janie Eubanks and Jeffery C. Valentine, 'Violent video games: Specific effects of violent content on aggressive thoughts and behavior' in Zanna M. P. (ed.), *Advances in Experimental Social Psychology: Vol. 36* (Elsevier Academic Press 2004).

[43] *Ibid*., 207 – 232.

behavior, affect, and cognition; on cardiovascular arousal; and on prosocial behavior', with more methodologically robust studies reporting stronger effects, contrary popular assertions.[44]

Both film and television are replete with advertising, and a natural line of inquiry asks whether adverts too might prime aggressive behaviours. Bartholow and Heinz present one study in which subjects were primed with advertisements of alcohol products, before completing a lexical decision-making task in the first experiment, and rating the behaviour of a target character in the second experiment.[45] Having been primed with alcohol advertisements, subjects were faster in accessing aggressive and hostile words in the lexical tasks, whilst they also rated target behaviours as being more aggressive in the behavioural-rating task. Predictably, an important moderator was the strength with which participants already associated alcohol with aggression. A similar study by Pedersen, Vasquez, Bartholow, Grosvenor and Truong also used alcohol advertisements to prime aggressive constructs.[46] After being primed, subjects in the first experiment increased the aggressiveness of their retaliation to an ambiguous provocation, whilst the second experiment further revealed the subjects' perception of the provocateur's hostility as a moderating factor.

A further study by Buchanan reveals that advertisements containing violent content on social media (specifically Facebook) can also prime higher levels of aggression-related cognition in subjects compared with non-violent adverts.[47] Relatedly, a number of studies have also demonstrated how sex-related constructs may similarly be primed through

[44] *Ibid*., 199 – 200.
[45] Bruce D. Bartholow and Adrienne Heinz, 'Alcohol and aggression without consumption: Alcohol cues, aggressive thoughts, and hostile perception bias' (2006) 17(1) *Psychological Science* 30.
[46] William C. Pedersen, Eduardo A. Vasquez, Bruce D. Bartholow, Marianne Grosvenor and Ana Truong, 'Are you insulting me? Exposure to alcohol primes increases aggression following ambiguous provocation' (2014) 40(8) *Personality and Social Psychology Bulletin* 1037.
[47] Tom Buchanan, 'Aggressive priming online: Facebook adverts can prime aggressive cognitions' (2015) 48 *Computers in Human Behavior* 323.

different media[48] – including videogames[49] and advertising[50] – such as the priming of gender stereotypes, sexually objectifying thoughts and sexually harassing behaviour. The emerging evidence suggests that sexualised content can prime both positive and negative sexual constructs; with regards to the latter, constructs such as regarding women as objects, perpetuating rape myths, and the normalisation of sexual violence may be primed, each of which can be important contributing factors towards criminal sexual behaviour and sexual violence.[51]

Finally with regards to aggression, as with research from television and film priming, there is growing evidence indicating an aggressive priming effect from violent videogames also,[52] similarly supported by a number of meta-analyses.[53] Equally a number of moderators have been illuminated, including initial trait aggression as

[48] Edward Donnerstein and Daniel Linz, 'Mass media sexual violence and male viewers' (1986) 29(5) *American Berhavioral Scientist* 601.

[49] Mike Z. Yao, Chad Mahood and Daniel Linz, 'Sexual priming, gender stereotyping, and likelihood to sexually harass: Examining the cognitive effects of playing a sexually-explicit video game' (2010) 62(1/2) *Sex Roles* 77.

[50] Christine Hall Hansen and Walter Krygowski, 'Arousal-augmented priming effects: Rock music videos and sex object schemas' (1994) 21(1) *Communication Research* 24; Francesca R. Dillman Carpentier, C. Temple Northup and M. Scott Parrott, 'Revisiting media priming effects of sexual depictions: Replication, extension, and consideration of sexual depiction strength' (2014) 17(1) *Media Psychology* 34.

[51] See further Michael L. Capella, Ronald Paul Hill, Justine M. Rapp and Jeremy Kees, 'The impact of violence against women in advertisements' (2010) 39(4) *Journal of Advertising* 37; Francesca R. Dillman Carpentier, 'Priming' in Rössler P., Hoffner C. A. and van Zoonen L. (eds.), *The International Encyclopedia of Media Effects* (John Wiley & Sons 2017), 11 – 12; John Davies, He Zhu and Brian Brantley, 'Sex appeals that appeal: Negative sexual self-schema as a moderator of the priming effects of sexual ads on accessibility' (2007) 29(2) *Journal of Current Issues & Research in Advertising* 79.

[52] For example, see further Mary E. Ballard and J. Rose West, 'Mortal Kombat™: The effects of violent videogame play on males' hostility and cardiovascular responding' (1996) 26(8) *Journal of Applied Social Psychology* 717; Craig A. Anderson and Karen E. Dill, 'Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life' (2000) 78(4) *Journal of Personality and Social Psychology* 772; Cameron D. Panee and Mary E. Ballard, 'High versus low aggressive priming during videogame training: Effects on violent action during game play, hostility, heart rate, and blood pressure' (2002) 32(12) *Journal of Applied Social Psychology* 2458; Wolfgang Bösche, 'Violent video games prime both aggressive and positive cognitions' (2010) 22(4) *Journal of Media Psychology* 139.

[53] For example, Craig A. Anderson and Brad J. Bushman, 'Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior: A meta-analytic review of the scientific literature' (2001) 12(5) *Psychological Science* 353; Craig A. Anderson, 'An update on the effects of playing violent video games' (2004) 27(1) *Journal of Adolescence* 113; Craig A. Anderson, Akiko Shibuya, Nobuko Ihori, Edward L. Swing, Brad J. Bushman, Akira Sakamoto, Hannah R. Rothstein and Muniba Saleem, 'Violent video games effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review' (2010) 136(2) *Psychological Bulletin* 151; Anna T. Prescott, James D. Sargent and Jay G. Hull, 'Meta-analysis of the relationship between violent video game play and physical aggression over time' (2018) 115(40) *Proceedings of the National Academy of Sciences* 9882.

previously discussed, as well as new potential moderators such as competitiveness, difficulty and pace of action in the violent videogames.[54] However, there is greater suspicion surrounding videogame priming and some analyses have failed to find similar effects;[55] even proponents concede that the reported effects are smaller than those in relation to film and television.[56] This uncertainty within the evidence likely results from the interplay of moderators of the priming effect with different experimental designs.[57] There may be a range of reasons why videogames might produce weaker priming effects, from less realism in the animated scenes of videogames compared to film and television scenes, to greater variation in the arousal this different media form produces in subjects. A number of recent and particularly revealing experiments in this regard have shown that the aggression priming effect from videogames is stronger when subjects can personalise their playable character or avatars, thus creating a stronger conceptual link between themselves and the gameplay.[58]

The vast majority of studies considered above concerning priming aggression are discussed within the context of Allen and Anderson's "General Aggression Model" of aggressive behaviour that has been applied to the understanding of a wide range of

[54] Craig A. Anderson and Nicholas L. Carnagey, 'Causal effects of violent sports video games on aggression? Is it competitiveness of violent content?' (2009) 45(4) *Journal of Experimental Social Psychology* 731; Paul J. C. Adachi and Teena Willoughby, 'The effect of violent video games on aggression: Is it more than just the violence?' (2011) 16(1) *Aggression and Violent Behavior* 55.

[55] Derek Scott, 'The effect of video games on feelings of aggression' (1995) 129(2) *Journal of Psychology* 121; Mark Griffiths, 'Violent video games and aggression: A review of the literature' (1999) 4(2) *Aggression and Violent Behavior* 203; David Zendle, Paul Cairns and Daniel Kudenko, 'No priming in video games' (2017) 78 *Computers in Human Behavior* 113.

[56] See John L. Sherry, 'The effects of violent video games on aggression: A meta-analysis' (2006) 27(3) *Human Communications Research* 409.

[57] In this regard (with accompanying meta-analysis), see John L. Sherry, 'Violent video games and aggression: Why can't we find effects?' in Preiss R. W., Gayle B. M., Burrell N., Allen M. and Bryant J. (eds.), *Mass Media Effects Research: Advances Through Meta-Analysis* (Routledge 2007); Douglas A. Gentile and Craig A. Anderson, 'Violent video games: The newest media violence hazard' in Gentile D. A. (ed.), *Advanced in Applied Developmental Psychology. Media Violence and Children: A Complete Guide for Parents and Professionals* (Praeger Publishers 2003).

[58] Peter Fischer, Andreas Kastenmüller and Tobias Greitmeyer, 'Media violence and the self: The impact of personalized gaming characters in aggressive video games on aggressive behavior' (2009) 46(1) *Journal of Experimental Social Psychology* 192; Jorge Peña, Jeffrey T. Hancock and Nicholas A. Merola, 'The priming effects of avatars in virtual settings' (2009) 36(6) *Communication Research* 838; Jorge Peña, Matthew S. McGlone and Joseph Sanchez, 'The cowl makes the monk: How avatar appearance and role labels affect cognition in virtual worlds' (2012) 5(3) *Journal for Virtual Worlds Research* 1; Grace S. Yang, L. Rowell Huesmann and Brad J. Bushman, 'Effects of playing a violent video game as male versus female avatar on subsequent aggression in male and female players' (2014) 40(6) *Aggressive Behavior* 537.

common violent behaviours.[59] First, input variables (which may relate to an individual's personality or a situation in which they find themselves) can influence affect (*i.e.,* mood and emotions) and cognitions (*i.e.,* thoughts) and can increase or decrease physiological and psychological arousal. Second, cognitive concepts (such as are linked to aggression) can be made accessible by priming with aggression, for example through exposure to media violence. Thus, priming may operate as an input variable that increases arousal, which can in turn influence aggressive behaviour in three ways. First, dominant tendencies towards aggression can be stimulated by irrelevant sources (primes); 'if an individual happens to be provoked while already in a state of high arousal, aggressive action tendencies can be strengthened.'[60] Second, such arousal from irrelevant sources might be misattributed anger, further encouraging an aggressive behavioural response. Third, abnormally high or low arousal (such as might be induced by priming effects) 'may be unpleasant states that encourage aggression in the same way that high temperatures or physical discomfort do.'[61] Consequently, whereas priming alone is unlikely to necessarily induce an individual into overt aggression or violence, it remains nonetheless appreciable how primes can be one important causative factor amongst many in instances of violent behaviour.

*

Another broad category of behavioural responses that may readily precursor criminal conduct are those relating to acting dishonestly or cheating. Acting in such a way that is "dishonest" is a crucial *mens rea* component of the most commonly prosecuted criminal offence – theft – as well as a range of other (largely property) offences. As aggressive responses may be primed, it should be unsurprising that dishonest responses may

[59] See generally, Craig A. Anderson and Nicholas L. Carnagey, 'Violent evil and the general aggression model' in Miller A. G. (ed.), *The Social Psychology of Good and Evil* (1st ed. The Guildford Press 2004); Craig A. Anderson and L. Rowell Huesmann, 'Human aggression: A social-cognitive view' in Hogg M. A. and Cooper J. (eds.), *The SAGE Handbook of Social Psychology* (SAGE Publications 2007); C. Nathan DeWall, Craig A. Anderson and Brad J. Bushman, 'Aggression' in Tennen H., Suls J. and Weiner I. B. (eds.), *Handbook of Psychology: Vol 5* (2nd ed. John Wiley & Sons 2012).

[60] Johnie J. Allen and Craig A. Anderson, 'General aggression model' in Rössler P. and Hoffner C. A. (eds.), *The International Encyclopedia of Media Effects* (John Wiley & Sons 2017), 10.

[61] *Ibid*.

similarly be primed.[62] In a series of experiments, Gino and Ariely[63] first demonstrated that people with more creative personalities tended to cheat more than less creative people, following the theory that they are correspondingly better able to justify their behaviour. They proceeded to show that subjects primed to think creatively were in turn more likely to behave dishonestly than controls, and possessed a correspondingly greater ability to provide justifications for their behaviour. In an earlier experiment, Gino and Pierce revealed how subjects primed with visible proximity to wealth – including by being close to an ostensibly wealthy person – resulted in more frequent cheating by overstating their performance on a subsequent anagram task.[64]

DeBono, Shariff, Poole and Muraven also offer three experiments which demonstrate that priming the idea of a forgiving god (but *not* a punishing god) amongst Christian subjects resulted in increased unethical behaviour.[65] In this experiment, the fact that all subjects were Christian is almost certainly a relevant moderator for using religious primes. Other cultural moderators have emerged through exploring priming in banking. In an initial experiment by Cohn, Fehr and Maréchal,[66] employees within the banking industry were first primed with the idea of "banking culture" by being asked to reflect on their own employment and professional background. Subjects subsequently performed a coin-tossing task in which they had the opportunity to cheat and misreport their results in order to win more money. The experiment showed that those primed with banking culture performed significantly more dishonestly than control subjects who had not been so primed.[67] However, a much larger replication across five different populations from three continents produced inconsistent results, finding notable variation across the different

[62] See Stephen Mark Rosenbaum, Stephan Billinger and Nils Stieglitz, 'Let's be honest: A review of experimental evidence of honesty and truth-telling' (2014) 45 *Journal of Economic Psychology* 181, 189.
[63] Francesca Gino and Dan Ariely, 'The dark side of creativity: Original thinkers can be more dishonest' (2012) 102(3) *Journal of Personality and Social Psychology* 445.
[64] Francesca Gino and Lamar Pierce, 'The abundance effect: Unethical behavior in the presence of wealth' (2009) 109(2) *Organizational Behavior and Human Decision Processes* 142.
[65] Amber DeBono, Azim F. Shariff, Sarah Poole and Mark Muraven, 'Forgive us our trespasses: Priming a forgiving (but not a punishing) God increases unethical behavior' (2017) 9(Supp 1) *Psychology of Religion and Spirituality* S1.
[66] Alain Cohn, Ernst Fehr and Michel André Maréchal, 'Business culture and dishonesty in the banking industry' (2014) 516(7529) *Nature* 86.
[67] *Ibid*., 87.

populations.[68] The authors of this latter study suggest that the inconsistency reflects variations in national banking cultures, industry segments, and norms of honesty and dishonesty, *etc*.[69] This demonstrates how, much like with religious primes operating on people who hold religious beliefs, priming certain ideas from the notion of "banking culture" will vary according to the different cultural views and stereotypes that are held regarding that industry.[70]

Contrary to much of the discussion in this chapter, however, it is not at all the case that priming can only influence negative behaviours. Within the area of dishonesty and cheating, for example, numerous experiments demonstrate how priming concepts of honesty can also decrease later cheating behaviour.[71] Evidence continues to emerge regarding the different factors that can moderate priming of dishonest or cheating behaviour. For example, one obvious moderating factor is the degree to which an individual possesses initial trait honesty or dishonesty, with those subjects scoring higher levels of trait dishonesty being more likely to act dishonestly in response to priming,[72] mirroring the relationship between trait hostility and response to aggression primes. In a similar vein, where emotional disgust has been demonstrated to prime dishonest behaviour, this effect is itself moderated by individual sensitivity to disgust.[73]

*

[68] Zoe Rahwan, Erez Yoeli and Barbara Fasolo, 'Heterogeneity in banker culture and its influence on dishonesty' (2019) 575(7782) *Nature* 345.
[69] *Ibid*., 349.
[70] See further Alain Cohn and Michel André Maréchal, 'Priming in economics' (University of Zurich Department of Economics, Working paper series no. 226, 2016), 2 – 4.
[71] For example, see Yu-Wei Wu, Lu-Lu Zhong, Qian-Nan Ruan, Jing Liang and Wen-Jing Yan, 'Can priming legal consequences and the concept of honesty decrease cheating during examinations?' (2020) 10 *Frontiers in Psychology* 2887; Aaron D. Nichols, Martin Lang, Christopher Kavanagh, Radek Kundt, Junko Yamada, Dan Ariely and Panagiotis Mitkidis, 'Replicating and extending the effects of auditory religious cues on dishonest behavior' (2020) 15(8) *PLoS ONE* e0237007; Mark E. Aveyard, 'A call to honesty: Extending religious priming of moral behavior to Middle Eastern Muslims' (2014) 9(7) *PLoS ONE* e99447.
[72] Emmanuel P. Kleinlogel, Joerg Dietz and John Antonakis, 'Lucky, competent, or just a cheat? Interactive effects of honesty-humility and moral cues on cheating behavior' (2017) 44(2) *Personality and Social Psychology Bulletin* 158.
[73] How Hwee Ong, O'Dhaniel A. Mullette-Gillman, Kenneth Kwok and Julian Lim, 'Moral judgment modulation by disgust is bi-directionally moderated by individual sensitivity' (2014) 5 *Frontiers in Psychology* 194; Julian Lim, Paul M. Ho and O'Dhaniel A. Mullette-Gillman, 'Modulation of incentivized dishonesty by disgust facial expressions' (2015) 9 *Frontiers in Neuroscience* 250.

Finally to consider in this section are a number of experiments revealing how the impressions that people form about others and their behaviours can again be primed in a particular direction. Impression formation is a critical aspect of social interactions, and the impressions that people form regarding others and their behaviour influence in turn how people respond to one another in social situations. When confronted with behaviour that is interpreted as hostile or aggressive, for example, people are more likely to respond in kind. Equally, when people gain the impression that others are cheating or acting dishonestly, they are more likely to respond in a similar manner. If the very impressions that we form about others can also be primed, therefore, it follows that our own responsive behaviour may be based on the indirect influence of those primes and not necessarily following reason or conscious reflection. Fiske and Neuberg further discuss how the relative accessibility of social categories – which may be increased by primes – is an important component contributing to how people perceive the actions and behaviour of others.[74]

Introduced in section 3.1, above, Higgins, Rholes and Jones[75] provide an original paradigm which has since been replicated robustly across a range of contexts.[76] Subjects were first primed with particular personality traits in what they believe to be an initial experiment regarding perception. In each trial subjects were first given a memory word – the priming words were embedded within these memory words – followed by a task in which they had to name a presented colour, following which they had to repeat the memory word. Various different priming methods have since been used such as embedding priming words within a wordsearch task; the common feature is that semantic priming is used whereby words prime semantically related constructs in the mind.

[74] Susan T. Fiske and Steven L. Neuberg, 'A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation' (1990) 23 *Advances in Experimental Psychology* 1, 11 – 12.
[75] Higgins, Rholes and Jones (1977).
[76] Srull and Wyer (1979); Bargh, Bond, Lombardi and Tota (1986); John A. Bargh, Wendy J. Lombardi and E. Tory Higgins, 'Automaticity of chronically accessible constructs in person x situation effects on person perception: It's just a matter of time' (1988) 55(4) *Journal of Personality and Social Psychology* 599; Steven J. Sherman, Diane M. Mackie and Denise M. Driscoll, 'Priming and the differential use of dimensions in evaluation' (1990) 16(3) *Personality and Social Psychology Bulletin* 405; Thomas E. Ford and Arie W. Kruglanski, 'Effects of epistemic motivations on the use of accessible constructs in social judgment' (1995) 21(9) *Personality and Social Psychology Bulletin* 950; Melissa J. Ferguson, John A. Bargh and David A. Nayak, 'After-affects: How automatic evaluations influence the interpretation of subsequent, unrelated stimuli' (2005) 41(2) *Journal of Experimental Social Psychology* 182.

Following the "perception" experiment, subjects completed a reading comprehension study in which they were given a paragraph regarding a fictional character, describing activities in neutral terms which could be interpreted in different ways. For example, the character might be described as refusing to answer the door to a salesman, which could be interpreted as being hostile and unfriendly or, alternatively, could be perceived more reasonably. In each of the experiments cited, the vignettes were pilot tested to ensure that they produce a neutral and ambiguous reflection of the character described.

Finally, subjects answered a number of questions regarding their impression of the personality and behaviour of the ambiguously described character, typically responding on a Likert scale. The results of these experiments demonstrate that when presented with ambiguously described characters and actions, subjects are significantly more likely to interpret behaviour according to the personality trait with which they have been primed. Bargh makes two important observations concerning these experiments. First, the nature of the priming tasks are such that subjects are conscious and aware of the priming material; however, the experiments are carefully designed such that the primes are not so explicit that subjects become consciously aware of their connection with or influence over the subsequent impression formation tasks.[77] Second, he submits that the experiments provide evidence that 'categorization of social behaviors in terms of trait concepts is an automatic process' – *i.e.,* people interpret and categorise the behaviour of others automatically and not necessarily consciously.[78] This aspect of automaticity is expanded upon in the discussion on theories of priming in section 3.1.3 of this thesis, below.

### *3.1.2. Priming Goals Outside of awareness*

Where the previous section has demonstrated how a whole range of behavioural responses may be primed in an equally vast range of different circumstances, this section proceeds to consider the priming of goals and intentions. Responses and goals are qualitatively distinct; "responses" are, by nature, responses *to* something else, *i.e.,* an object, an

[77] Bargh and Chartrand (2014), 317.
[78] John A. Bargh, 'Automatic information processing: Implications for communication and affect' in Donohew L., Sypher H. E. and Higgins E. T. (eds.), *Communication, Social Cognition, and Affect* (Psychology Press 1988), 14 – 15.

environment, or some other stimuli. Conversely, goals or intentions operate *towards* some end state; they do not *necessarily* arise simply in response to some stimulus, and typically require some positive action or engagement in order to be realised. In this sense, responses and goals are often colloquially distinguished on account of responses (or reactions) being capable of operating quite automatically (as demonstrated in the previous section of this chapter, above), whereas goals and intentions are more typically regarded as requiring conscious volition or effort in order to be achieved. The goals which people choose to pursue might therefore often be regarded as proper or more true reflections of a person's character, personality and intentions, being an expression of their conscious will. However, the 'same higher mental processes that have traditionally served as quintessential examples of choice and free will – such as goal pursuit, judgment, and interpersonal behavior' – are revealed in this section to be capable of occurring 'in the absence of conscious choice or guidance.'[79]

A number of initial experiments demonstrate some of the crucial differences between priming goals as contrasted with other forms of priming. Hamilton, Katz and Leirer[80] provide an important set of original paradigms in which subjects were explicitly instructed with the goal of either forming an impression of a target character or remembering as much information as possible about that target. Counterintuitively at the time, those subjects primed with the goal of forming an impression of the target character later remembered more information about them, an effect which is explained by the involvement of information integration and organisation in forming an impression, which facilitates later retrieval of individual items of that information.[81] However, the "priming" in this experiment was in the form of an explicit instruction the purpose of which the subjects were clearly consciously aware.

[79] John A. Bargh and Melissa J. Ferguson, 'Beyond behaviorism: On the automaticity of higher mental processes' (2000) 126(6) *Psychological Bulletin* 925, 926.

[80] David L. Hamilton, Lawrence B. Katz and Von O. Leirer, 'Organizational processes in impression formation' in Hastie R., Ostrom T. M., Ebbesen E. B., Wyer R. S., Hamilton D. L. and Carlston D. E. (eds.), *Person Memory: The Cognitive Basis of Social Perception* (Lawrence Erlbaum Associates 1980a); David L. Hamilton, Lawrence B. Katz and Von O. Leirer, 'Cognitive representation of personality impressions: Organizational processes in first impression formation' (1980b) 39(6) *Journal of Personality and Social Psychology* 1050.

[81] Hamilton, Katz and Leirer (1980b), 1061 – 1062.

Chartrand and Bargh[82] replicated these earlier paradigms with a crucial change; instead of instructing subjects explicitly to pursue the impression formation or memorisation goal, subjects were primed with these goals outside of conscious awareness. Chartrand and Bargh first administered a scrambled-sentences priming paradigm in which subjects must complete grammatically correct four-word sentences out of five words presented in a scrambled order.[83] The sentences included words related to the concepts of evaluation / personality / impression, and memory / retaining / remembering, in order to prime the goals of impression formation and memorisation respectively. After a short filler task, subjects then completed a replication of the task from Hamilton, Katz and Leirer in which a number of sentences described actions or features of a target character, following which subjects were tested for how much of that information could be recalled. Chartrand and Bargh's finding closely paralleled the original pattern of results with subjects primed to form an impression of the target character remembering more information about them. Crucially, it showed that the 'information-processing goals that have been shown in previous work to produce differential organization and memory for social information when operating consciously and intentionally *have the identical effects on processing when operating automatically*.'[84]

Chartrand and Bargh replicated a further experiment by Bargh and Thein,[85] again with a critical change in order to prime subjects outside of their conscious awareness. Subjects were first primed with the goal of forming an impression of a target character through a parafoveal vigilance task during which words related or unrelated to impression formation were presented briefly away from a fixation point at which subjects stared. Next, subjects were presented with a series of descriptions of honest, dishonest and neutral behaviour of a target character, followed by a brief filler task. Finally, a surprise free-recall task ascertained how many behaviour descriptions subjects could remember, following which subjects were asked to report their overall impressions of the character

[82] Tanya L. Chartrand and John A. Bargh, 'Automatic activation of impression formation and memorization goals: Nonconscious goal priming reproduces effect of explicit task instructions' (1996) 71(3) *Journal of Personality and Social Psychology* 464.
[83] See Srull and Wyer (1979).
[84] Chartrand and Bargh (1996), 469.
[85] John A. Bargh and Roman D. Thein, 'Individual construct accessibility, person memory, and the recall-judgment link: The case of information overload' (1985) 49(5) *Journal of Personality and Social Psychology* 1129.

described. In contrast to subjects with no priming, those primed with the goal of forming an impression of the target were significantly more likely to form an on-line impression of the character whilst reading the description *and* recall more details about that character later on.[86]

Taken together, the experiments duly replicated the results of previously established paradigms whilst providing subjects with their goal via priming outside of awareness as opposed to explicit instruction. The authors interpret these results as strongly supporting the contention that 'intentions and goals can be automated and that their effects when operating nonconsciously are identical to their effects when they are operating consciously and deliberately.'[87] A related experiment by Woike, Lavezzary and Barsky[88] further illustrated that the implicit cognitive consequences of goal processing are the same for goals that are implicitly accessible due to a chronic state of the subject or due to a temporary triggering event. Subsequently, Bargh, Lee-Chai, Barndollar, Gollwitzer and Trötschel[89] conducted a further five experiments with the aim of demonstrating that goals primed implicitly outside of awareness could process and execute with all the same features of ordinary explicit goal pursuit.[90] Such "classic features" of ordinary goal pursuit identified by Bargh *et. al.* include 'vigorous acting toward goal attainment, persistence in the face of obstacles, and resumption after disruption.'[91]

In the first experiment, subjects were primed using a wordsearch puzzle with words relating to the goal of high-performance such as "success", "winning" and "competition", following which their performance was measured on three further experimental wordsearch puzzles. Finally, a funnelled questionnaire protocol checked that subjects did not suspect the purposes of, or relationship between, the different tasks, ensuring that the

[86] Chartrand and Bargh (1996), 472.
[87] *Ibid*., 475.
[88] Barbara Woike, Erica Lavezzary and Jennifer Barsky, 'The influence of implicit motives on memory processes' (2001) 81(5) *Journal of Personality and Social Psychology* 935.
[89] John A. Bargh, Annette Lee-Chai, Kimberly Barndollar, Peter M. Gollwitzer and Roman Trötschel, 'The automated will: Nonconscious activation and pursuit of behavioral goals' (2001) 81(6) *Journal of Personality and Social Psychology* 1014.
[90] *Ibid*., 1016.
[91] *Ibid*., 1018; citing Robert A. Wicklund and Peter M. Gollwitzer, *Symbolic Self-Completion* (Routledge 1982); Heinz Heckhausen, *Motivation and Action* (Springer New York 1991).

high-performance goal was primed outside of subjects' conscious awareness. The results successfully showed that performance on the experimental wordsearches was enhanced when subjects had been primed with the goal of high performance relative to those subjects who had not been so primed. This established that 'performance goals can become activated without the necessity of conscious and deliberate choice and then operate to regulate behaviour towards attainment of the desired outcome.'[92]

In the second experiment by Bargh *et. al.*, subjects were first primed with the goal of cooperation using the scrambled sentences paradigm from Srull and Wyer, described previously in this chapter. This was followed by a resource-management game in which subjects could either play to make personal profit but deplete a common resource, play to cooperate and maintain that resource, or apply some middling strategy between the two extremes. Some subjects were further given explicit instructions to cooperate, thereby providing an explicit goal condition to compare against the goal primed unconsciously. The results revealed the significant effect of both implicit (unconscious) and explicit (conscious) goals to cooperate, with the latter being stronger than the former. Furthermore, comparison of the primed versus instructed subjects confirmed that 'nonconscious goal activation does not require the pre-existence of a conscious goal in the same direction (*i.e.,* the piggybacking issue).'[93] A third experiment used a dissociation paradigm in order to exclude alternative explanations for the results of the first two experiments.

Experiment four investigated the persistence of nonconsciously primed goals when confronted with obstacles. Subjects were again first primed with stimuli relating to high performance before being given wordsearch tasks similar to experiment one. However, rather than being permitted the full time to complete the wordsearch, subjects were interrupted after two minutes, thus preventing satisfaction of the primed high-performance goal. The dependent measure was whether or not subjects continued to proceed with the task to achieve a higher score in spite of the stop signal, whilst surreptitiously being viewed on a hidden camera. The results showed that a significantly

[92] Bargh *et. al.* (2001), 1021.
[93] *Ibid*., 1023.

greater number of subjects continued with the wordsearch task beyond the stop signal having been primed with the high-performance goal.[94]

Finally, the fifth experiment from Bargh *et. al*. explored the resumption of primed goals that had been interrupted. Subjects were primed with the high-performance goal using the wordsearch paradigm from the previous experiments. This was followed by a word-construction task in which subjects had to construct as many words as possible from scrambled letters. After only one minute, however, this intellectual task was interrupted by a purported technical fault taking some 5 minutes to fix. At this point, subjects were told that there was insufficient time to complete the first intellectual task properly, and that they could choose between either continuing this task regardless or switching to a final task in which they rated different cartoons for their humour. This final task was intended to entice subjects away from the less fun intellectual task. Here, again, subjects that had been primed nonconsciously with the goal of high performance were 'considerably more likely to return to the incomplete intellectual task after interruption than were nonprime participants.'[95]

On the one hand, these initial experiments demonstrate how goals that have been primed outside of conscious awareness appear to operate, process and execute identically to ordinary explicit and conscious goals, which is to say that consciousness itself does not appear to be a prerequisite of successful goal activation and pursuit. On the other hand, these experiments also demonstrate critical differences between priming goals and other forms of semantic and conceptual priming relating to perceptions and behavioural responses. Whereas other forms of priming tend to be temporary, short-lived and subtle in their effect, the strength of an unconsciously primed goal 'looms larger as time passes and it remains unfulfilled.'[96] Also contrary to semantic and response priming, people are more likely to persist with primed goals when confronted with obstacles and, further still,

[94] *Ibid*., 1030.
[95] *Ibid*., 1032.
[96] Gordon B. Moskowitz, Peizhong Li and Elizabeth R. Kirk, 'The implicit volition model: On the preconscious regulation of temporarily adopted goals' in Zanna M. P. (ed.), *Advances in Experimental Social Psychology, Vol. 36* (Elsevier Academic Press 2004), 335; see also Nira Liberman and Jens Förster, 'Expression after suppression: A motivational explanation of postsupressional rebound' (2000) 79(2) *Journal of Personality and Social Psychology* 190.

to resume the pursuit of primed goals after being interrupted. These differences identify the priming of goals and intentions as being crucially important not only for their distinctiveness from other more subtle forms of priming, but also for the potential of such priming effects to significantly impact upon subsequent decisions and behaviour, all the while outside of the explicit conscious awareness of the individual subject.

*

A set of experiments by Aarts and Dijksterhuis[97] not only offer a further display of priming goals but also reveal how plans of action can become automatically associated with the goals that they are intended to accomplish, thus demonstrating unconscious processing from the priming of a goal through to the preparation of its relevant implementing action. The first experiment investigated the hypothesis that habitual actions would be automatically activated upon the instigation of a relevant goal whilst similar actions would not be activated amongst those for whom that behaviour was not habitual.[98] Subjects were habitual and non-habitual cyclists who either did or did not regularly use a bicycle to travel to university. For the priming task, subjects were asked to read sentences and press a button after reading each one in what they believed was a measure of reading speed; amongst these priming stimuli were sentences describing different travel goals to specific locations, such as "going shopping at the city centre mall". Subjects subsequently performed an association task in which different location words were presented followed by a mode of transport, and the subjects had to indicate as quickly as possible whether the presented mode of transport was a realistic means of travelling to the presented location.

The results showed that subjects who habitually cycled were significantly faster in responding to target-location pairs involving use of a bicycle when they had also been primed with the travelling goal, suggesting that the 'automaticity of habitual behaviors is

[97] Henk Aarts and Ap Dijksterhuis, 'Habits as knowledge structures: Automaticity in goal-directed behavior' (2000a) 78(1) *Journal of Personality and Social Psychology* 53; Henk Aarts and Ap Dijksterhuis, 'The automatic activation of goal-directed behaviour: The case of travel habit' (2000b) 20(1) *Journal of Environmental Psychology* 75.
[98] Aarts and Dijksterhuis (2000a), 55.

conditional on the presence of a goal' which itself may be exogenously primed outside of conscious awareness.[99] These findings were supported by two further experiments in the same study by Aarts and Dijksterhuis. Experiment two replicated the results of experiment one whilst further revealing how planning to use a bicycle made non-habitual users faster at identifying correct associations between target-location pairs involving a bicycle, whilst habitual users obtained no comparable benefit through planning. This was interpreted to suggest that habits are 'mentally represented as associations between goals and actions.'[100] The third experiment further clarified that the 'automatic activation of a habitual response is conditional on the presence of a travel goal.'[101] Read together, the experiments by Aarts and Dijksterhuis suggest that habitual behaviours may be automatically linked to mental representations of the goals that they have developed to serve and, as such, may be activated outside of conscious awareness by priming those relevant goals.

Exploring the question of what may operate to prime goals in a more real-world context, Aarts, Gollwitzer and Hassin[102] present a series of experiments demonstrating the goal contagion hypothesis which claims that people may automatically adopt and pursue – *i.e.,* be primed by – goals that are implied from the behaviour of others.[103] This builds upon earlier work showing that people are readily able to infer goals from observing others' behaviour and,[104] moreover, that such goal inferences can be made automatically and outside of conscious awareness.[105] In the first experiment, subjects first read a short story about a character which served either as a control, or described behaviour which was confirmed in a pilot study to evoke the goal of earning money. This was followed by a second box-ticking task for which the subjects were informed they could earn additional

[99] *Ibid.*, 56 – 57.
[100] *Ibid.*, 59.
[101] *Ibid.*, 60.
[102] Hank Aarts, Peter M. Gollwitzer and Ran R. Hassin, 'Goal contagion: Perceiving is for pursuing' (2004) 87(1) *Journal of Personality and Social Psychology* 23.
[103] See also Giel Dik and Henk Aarts, 'Behavioral cues to others' motivation and goal pursuits: The perception of effort facilitates goal inference and contagion' (2007) 43(5) *Journal of Experimental Social Psychology* 727.
[104] Andrew N. Meltzoff and M. Keith Moore, 'Infants' understanding of people and things: From body imitation to folk psychology' in Bermúdez J. L., Marcel A. J. and Eilan N. (eds.), *The Body and the Self* (Massachusetts Institute of Technology Press 1995).
[105] Ran R. Hassin, Henk Aarts and Melissa J. Ferguson, 'Automatic goal inferences' (2005) 41(2) *Journal of Experimental Social Psychology* 129.

money; however, they had to click to erase the instruction message first to access the additional task. The speed at which subjects closed this dialogue box served as a measure of the operationalised goal to earn money, whilst the speed of clicking boxes in the subsequent task served as a measure of the effort of goal pursuit. The primed goal alone produced only a slight increase in speed of accessing the money-earning task; however, those who expressed a greater need for money when later questioned showed a significant effect of goal contagion from the priming task. Equally, whilst primed subjects displayed somewhat greater effort in the money-earning task, again, those subjects who were primed and also had a greater need for money worked faster, demonstrating a goal contagion effect.[106]

The second experiment by Aarts, Gollwitzer and Hassin sought to conceptually replicate the first with an alternative goal of seeking casual sex. First, two pilot studies confirmed that a short story similar to experiment one successfully evoked the character's goal of seeking casual sex, and that offering to help women was perceived by male students as an appropriate method of achieving that goal. After the priming task subjects were asked to provide feedback on the previous experiment, which they were informed was created by a male or female undergraduate student; the number of words used and length of time devoted to providing feedback served as the measures of goal contagion. Here a significant interaction was demonstrated between the primed goal and the gender of the alleged experimenter; male subjects were primed by the goal implied from the behaviour of another with the goal of seeking casual sex, but this goal was only subsequently expressed by way of helping another and providing additional feedback when they thought that other person was a woman.[107]

The third experiment by Aarts, Gollwitzer and Hassin examined a particular typifying characteristic of goal pursuit – the persistent (and even increased) activation of goals over time. Specifically, if the results of the first two experiments were due to purely cognitive and non-motivational priming effects then those effects would be expected to be short-lived, whereas the persistence of such effects provides a clear indicator that motivational

[106] Aarts, Gollwitzer and Hassin (2004), 27.
[107] *Ibid*., 28 – 29.

*goals* are indeed being primed and activated. Experiment three repeated the procedures of experiment two, except some subjects were provided with a five-minute filler task between being primed and completing the feedback task. The experiment replicated the previous results, demonstrating a significant goal contagion effect; more importantly, that effect remained and did not appear to diminish over time, indicating that motivational goals were indeed being activated by the priming task.[108] Three further experiments in the same study by Aarts, Gollwitzer and Hassin proceeded to reveal some of the moderators of goal contagion; most pertinently, subjects 'do not automatically adopt goals when the observed goal pursuit is conducted in an unacceptable manner, because the goal will then be perceived as unattractive.'[109] In a similar vein, notable research has revealed that goal contagion is more likely to occur between people belonging to the same social groups, further revealing how the effect is moderated.[110]

Aarts, Custers and Veltkamp[111] pick up from the previous work by Aarts, Gollwitzer and Hassin in considering how and why people proceed to pursue goals in the absence of conscious will and the factors which moderate this effect. They focus in particular on affective valence, referring to the positive or negative valence that is automatically given to a stimulus as opposed to any conscious state of feeling or emotion, arguing that it is this affective valence which 'acts as a basic source in determining the motivation to pursue a goal.'[112] They begin with the proposition that people represent actions in terms of their effects or possible application to a particular goal, so that priming the representation of such goals provides an automatic reference point for directing action towards their attainment. However, they add that the 'goal concept is more likely to motivate people to pursue the goal if the concept is directly followed by positive affect' and, equally, people are less likely to pursue the goal when it is followed by negative affect. The 'affective valence signals that the accessible goal is worth pursuing and puts

[108] *Ibid*., 29 – 30.
[109] *Ibid*., 23.
[110] Chris Loersch, Hank Aarts, B. Keith Payne and Valerie E. Jefferis, 'The influence of social groups on goal contagion' (2008) 44(6) *Journal of Experimental Social Psychology* 1555.
[111] Henk Aarts, Ruud Custers and Martijn Veltkamp, 'Goal priming and the affective-motivational route to nonconscious goal pursuit' (2008) 26(5) *Social Cognition* 555.
[112] *Ibid*., 556.

people into a state of readiness for goal pursuit.'[113] This proposal is built upon prior research similarly demonstrating how affective processes 'can moderate decision making and behavior very quickly without reaching conscious awareness.'[114] Thus, positive affect following a primed goal creates the motivation to pursue that goal, whilst immediately following negative affect dampens any such motivation.

Aarts, Custers and Veltkamp present two original experiments in support of their proposal, taking advantage of the phenomenon that goals can bias perceptual processes such as by making objects that are more relevant to a particular goal appear to be visually larger than other non-relevant objects.[115] Subjects first completed a priming task in which they had to report whether a dot on a computer screen appeared above or below "irrelevant" words. These words in fact contained the primes; subjects were primed with neutral words related to the goal of completing puzzles, following which subjects were immediately presented with positive or neutral words, thus activating either positive or neutral affect immediately after priming the puzzle-completion goal. Subjects were then asked to estimate the size of two objects related to completing puzzles (a crossword puzzle and a puzzle book) either immediately or after a short two-and-a-half minute's delay. Subjects consequently estimated the size of the puzzle-relevant objects to be significantly larger in the positive affect condition, supporting the assertion that people are more ready to pursue a nonconsciously primed goal when that goal is accompanied by immediate positive affect.[116]

In a second experiment replicating the procedures of the first, subjects were additionally provided with a memorisation task which operated to increase cognitive load whilst also

---

[113] *Ibid*., 559 – 560; citing Ruud Custers and Henk Aarts, 'Beyond priming effects: The role of positive affect and discrepancies in implicit processes of motivation and goal pursuit' (2005a) 16(1) *European Review of Social Psychology* 257; Ruud Custers and Henk Aarts, 'Positive affect as implicit motivator: On the nonconscious operation of behavioral goals' (2005b) 89(2) *Journal of Personality and Social Psychology* 129; Ruud Custers and Henk Aarts, 'In search of the nonconscious sources of goal pursuit: Accessibility and positive affective valence of the goal state' (2007) 43(2) *Journal of Experimental Social Psychology* 312.

[114] Aarts, Custers and Veltkamp (2008), 560.

[115] See Claus Bundesen, Thomas Habekost and Soren Kyllingsbaek, 'A neural theory of visual attention: Bridging cognition and neurophysiology' (2005) 112(2) *Psychological Review* 291; John T. Serences and Steven Yantis, 'Selective visual attention and perceptual coherence' (2006) 10(1) *Trends in Cognitive Sciences* 38.

[116] Aarts, Custers and Veltkamp (2008), 567 – 568.

keeping the nonconscious goal active. The results of the first experiment were replicated and the neutral goal treated with positive affect persisted beyond the delay task. However, the priming effect disappeared for subjects given the additional memorisation task suggesting that their ability to keep the goal active was impaired by the secondary task competing for cognitive resources, 'thereby corroborating the suggestion that the nonconsciously shaped goal remained active by a kind of updating or rehearsal process.'[117] Together, these experiments not only further demonstrate the nonconscious priming of goals, but reveal why such primed goals may be pursued by individuals when they are accompanied with positive affect, again, operating outside of conscious awareness.

One final illuminating study to consider by Verbruggen and Logan begins to draw links between the automaticity of processing active goals with a similar potential for automaticity in processing inhibitive goals.[118] In the studies discussed thus far subjects have been primed with certain "active" goals, *i.e.,* goals to do a particular thing; Verbruggen and Logan therefore investigate whether it is similarly possible to prime "negative" goals, *i.e.,* goals to refrain from, or inhibit, doing a particular thing. The researchers adapted the stop-signal and go/no-go paradigms in which subjects must respond as quickly as possible to "go" or "stop" signals. The paradigms were manipulated across three experiments to investigate respectively whether subjects could be primed by both task-relevant and task-irrelevant information in order to "stop" on "go" trials.

Comparing results across the three experiments, the authors found highly consistent priming effects which 'clearly support the idea that cognitive control can be triggered in a stimulus-driven (unintentional) fashion as well as in a top-down (intentional) fashion.'[119] In a similar vein, research shows that the nonconscious activation of one goal (for example, studying) can also automatically inhibit competing goals (such as socialising), further suggesting towards the 'nonconscious operation of an inhibition

[117] *Ibid*., 570.
[118] Frederick Verbruggen and Gordon D. Logan, 'Automaticity of cognitive control: Goal priming in response-inhibition paradigms' (2009) 35(5) *Journal of Experimental Psychology* 1381.
[119] *Ibid*., 1385.

mechanism that shields goals from distracting thoughts.'[120] This theme of automatic or unconsciously driven inhibition / self-control is considered in greater detail under the *whether* dimension of decision-making in chapter six of this thesis, below.

*

A crucial element in the priming studies considered throughout this chapter is that subjects remain unaware of the specific influence or operation of the primes. Consequently, these experiments reveal the unconscious and automatic nature of the mental process involved; for example, in the experiments priming certain goals, those goals are subsequently activated and pursued automatically and outside of the subjects' conscious awareness, revealing how the underlying mental processes are not dependent upon conscious intervention or control. This point is explored in greater detail in the following section, below, considering theories of priming and automatic mental processes. As Bargh and Ferguson summarise:

> '[A]lthough the currently persuasive distinction in cognitive science between automatic and controlled mental processes makes it perhaps difficult to conceive of automatic control, we note that the term has been common in engineering for nearly 50 years and means the same thing there that we mean by it here: autonomous systems interacting with environmental information over time to attain a goal, without any need of intervention from outside that closed system to do so. *It is not necessary to invoke the idea of free will or a nondetermined version of consciousness as a causal explanatory mechanism in accounting for higher mental processes in humans.* '[121]

[120] Aarts, Custers and Veltkamp (2008), 558; citing James Y. Shah, Ron Friedman and Arie W. Kruglanski, 'Forgetting all else: On the antecedents and consequences of goal shielding' (2002) 83(6) *Journal of Personality and Social Psychology* 1261; Hans Marien, Ruud Custers, Ran R. Hassin and Henk Aarts, 'Unconscious goal activation and the hijacking of the executive function' (2012) 103(3) *Journal of Personality and Social Psychology* 399; Travis S. Crone, 'The effect of nonconscious goal conflict on goal-related behavior' (2016) 3(3) *Psychology of Consciousness: Theory, Research and Practice* 284.
[121] Bargh and Ferguson (2000), 939.

Finally, Moskowitz, Li and Kirk provide a valuable summation of the various interlinks between unconscious goal priming, volitional action, legal intention and self-control in a paragraph that is provided below. These concluding remarks also provide a convenient connection to discuss the underlying theories of priming and automatic behaviour.

> 'The debate surrounding the legal meaning of intent (and culpability for unintended effects arising from unconscious bias) must incorporate the issue of whether control can be automatic. It is posited that goal pursuit can be as dominant a response to a stimulus as the activation of other mental representations. Preconscious control is a common component of social life, it is just less intuitively obvious than conscious control (given lay conceptions of control and volition), more invisible or difficult to detect (given it is concealed by conscious rationalization), and more threatening to the individual's sense of identity (as it superficially suggests to the individual that action is determined by the environment alone). But implicit volition is a common component of everyday life, and it occurs even when the goals are not routinely practiced or associated with specific environments. We might use the term strategic automaticity to convey the possibility that temporary goals can be implicitly triggered and pursued – that implicit volition is not limited to chronic goals but can be selected by individuals within a current context to pursue context-relevant goals.'[122]

### *3.1.3. Theories of Priming*

It should be expected that priming effects 'vary by a wide range of moderating individual difference and experimental context variables.'[123] Further, whereas it might be reasonable to expect basic cognitive and perceptual priming effects to appear consistently across broad populations, the same cannot necessarily be said with regards to social priming effects. Approached from the widely held perspective of evolutionary psychology, the mind is a 'computational organ designed to incorporate information from different

[122] Moskowitz, Li and Kirk (2004), 404.
[123] Joseph Cesario, 'Priming, replication, and the hardest science' (2014) 9(1) *Perspectives on Psychological Science* 40, 41.

sources to regulate behaviour.'[124] In social interactions the brain must operate not only on information regarding the self and the environment, but on 'information beyond the target stimulus features' such as cultural knowledge and experiences, expectations and stereotypes held regarding the stimulus.[125] Crucially, such features readily vary between populations, cultures, and generations, *etc.*; 'in other words, more than just the primed stimulus information is needed to predict behaviour following priming and we should not expect responses to be broadly uniform.'[126] This does not necessarily undermine the power of priming or render the effect frivolous; it means that priming is more subtle, contextual, and sensitive to individual variations between different people.

This point may be exemplified by considering one of the seminal experiments by Bargh, Chen and Burrows in which subjects primed with African-American faces reportedly showed greater aggression towards an annoying request by the experimenter, discussed above in section 3.1.1. Since that experiment in 1996, however, subsequent research has revealed a range of variables which may specifically interact with priming aggressive responses to male African-American faces. For example, the extent to which an individual holds the stereotype that Black men are aggressive is a significant moderator of this particular priming effect.[127] It is understood that the aggressive behavioural responses to Black faces in the priming paradigm reflect the 'output of participants preparing to interact with a dangerous outgroup male,' such that if a person does not stereotype Black men as being aggressive then there is no need to 'prepare to interact with a dangerous

[124] *Ibid*., 43; citing John Tooby and Leda Cosmides, 'Conceptual foundations of evolutionary psychology' in Buss D. M. (ed.), *Handbook of Evolutionary Psychology* (John Wiley & Sons 2005); expanded in John Tooby and Leda Cosmides, 'The theoretical foundations of evolutionary psychology' in Buss D. M. (ed.), *Handbook of Evolutionary Psychology, Volume 1: Foundation* (John Wiley & Sons 2016).
[125] *Ibid*., 43; citing Joseph Cesario and Carlos David Navarrete, 'Perceptual bias in threat distance: The critical roles of in-group support and target evaluations in defensive threat regulation' (2013) 5(1) *Social Psychological and Personality Science* 12; Daniel M. T. Fessler and Colin Holbrook, 'Friends shrink foes: The presence of comrades decreases the envisioned physical formidability of an opponent' (2013) 24(5) *Psychological Science* 797.
[126] Cesario (2014), 44.
[127] Joseph Cesario, Jason E. Plaks, Nao Hagiwara and Carlos David Navarrete, 'The ecology of automaticity: How situational contingencies shape action semantics and social behavior' (2010) 21(9) *Psychological Science* 1311; Ap Dijksterhuis, Henk Aarts, John A. Bargh and Ad van Knippenberg, 'On the relation between associative strength and automatic behavior' (2000) 36(5) *Journal of Experimental Social Psychology* 531.

outgroup male.'[128] An individual's physical size is also a likely moderating factor following the argument that physically larger individuals can generally perform aggressive responses more effectively and at lower cost than smaller individuals.[129]

Simply the environment can also have an important moderating effect on priming responses, both in the experimental context and out in the world. Thus, White subjects seated in a small enclosed physical space had increased accessibility to "fight"-related semantics after being primed with Black faces, whilst subjects primed within an open physical environment accessed more "flight"-related semantics.[130] Similarly, subjects' being surrounded by other in-group members has been found to moderate priming effects following the principle that one can execute behaviours with the support of their coalition that could not as easily or safely be performed alone.[131] In addition to such specific moderators, there are other variables that have been shown to moderate priming effects more generally. For example, subjects who exhibit lower self-monitoring are more likely to be susceptible to priming effects than those exhibiting higher self-monitoring.[132] In a similar vein, the positive associations that an individual holds with regards to a particular prime has been shown to influence responses to that prime;[133] as has the association of a primed goal to in- or out-group members;[134] whilst the effect of primes has also been shown to be moderated according to whether they are generated by an individual themselves or by another.[135] Consequently, the change to some critical feature of an

[128] Cesario (2014), 42; citing Joseph Cesario, Jason E. Plaks and E. Tory Higgins, 'Automatic social behavior as motivated preparation to interact' (2006) 90(6) *Journal of Personality and Social Psychology* 893.
[129] Aaron Sell, John Tooby and Leda Cosmides, 'Formidability and the logic of human anger' (2009) 106(35) *Proceedings of the National Academy of Sciences* 15073.
[130] Cesario, Plaks, Hagiwara and Navarrete (2010).
[131] Fessler and Holbrook (2013).
[132] Kenneth G. DeMarree, S. Christian Wheeler and Richard E. Petty, 'Priming a new identity: Self-monitoring moderates the effects of nonself primes on self-judgments and behavior' (2005) 89(5) *Journal of Personality and Social Psychology* 657.
[133] Custers and Aarts (2005b).
[134] Loersch, Aarts, Payne and Jefferis (2008).
[135] Thomas Mussweiler and Roland Neumann, 'Sources of mental contamination: Comparing the effects of self-generated versus externally provided primes' (2000) 36(2) *Journal of Experimental Social Psychology* 194.

experimental design unguided by any underlying theoretical context may readily alter, reverse or even eliminate a reported priming effect within a different population.[136]

Consideration of theories of priming is particularly relevant for the theme of this first part of the thesis because each of the different theories share the common feature of describing automatic processes to explain how and why priming operates outside of conscious awareness. The importance of this is in revealing how the content of the *what* component of any decision can be initiated, processed and carried through into action, all entirely outside of a person's conscious awareness. This is not necessarily to the exclusion of conscious intervention – and the timing at which consciousness may intervene is considered in greater detail in chapter five of this thesis, below. Nevertheless, to translate briefly into more legal parlance, the implications of the various nonconscious priming studies and underlying explanatory theories suggest that people may find themselves possessing the requisite *mens rea* for a particular offence – *i.e.,* some intention, dishonesty, or criminal knowledge etc. – without ever having consciously determined to pursue that criminal goal. That is to say, it is perfectly possible that some criminal goal or intention could be activated in the mind and, in the right circumstances, be processed through to the execution of criminal action without that individual's conscious awareness or realistic opportunity for conscious intervention.

Where the legal concept of *mens rea* is concerned solely with the content of a defendant's subjective mind, it can make no meaningful distinction between those defendants whose behaviour may have arisen largely automatically and outside of any possibility for conscious control, and those whose behaviour was consciously deliberated and determined to bring about a criminal act. The experiments considered above, alongside the theories of priming discussed below, reveal just how such criminal goals or intentions can (albeit not necessarily must) be processed in the mind from genesis to completion without the opportunity for conscious awareness or intervention. If this is accepted, then the focus of *mens rea* solely upon the subjective content of a defendant's mind fails to pay due regard to the origins and processing of that subjective content and, most crucially, may ignore entirely the absence of any realistic opportunity for intervention or control.

[136] Cesario (2014), 43.

Furthermore, the law places significance upon *mens rea* following the principle that a particular intended action was freely and deliberately chosen by an individual, whereas the fact that both behavioural responses and more complex goals may be triggered and acted upon entirely through automatic processes places a great challenge on this legal presumption.

*

The earliest and arguably most generalised theories of priming emerged from the early work of Higgins *et. al.* and Srull and Wyer, investigating the priming of social impressions regarding ambiguously described behaviour. As Molden describes, the initial mechanisms proposed to explain priming effects consisted of two components: '(1) the "excitation" of representations in memory by some process of spreading activation through a semantic network of associations, and (2) the use of these excited, or *accessible*, representations to encode information about a social target that was subsequently received.'[137] However, spreading activation alone cannot account for the many and varied interactions of different factors which moderate priming effects and necessarily require additional processes.[138] Furthermore, spreading activation can only adequately explain short-term priming effects and not the longer effects which have also been revealed (particularly in goal priming), unless it is assumed that a primed concept remains a source of activation beyond the priming stimulus itself.[139] Thus, the spreading of activation through semantically associated networks in the brain cannot account for the full range of priming effects. Nonetheless, this concept of spreading activation, which occurs automatically and without conscious intervention, remains a common *initial component* of many subsequent theories of priming.

The ground-breaking work of Bargh and Dijksterhuis – demonstrating how priming could not only impact automatically upon impressions and perceptions but also social

[137] Daniel C. Molden, 'Understanding priming effects in social psychology: What is "social priming" and how does it occur?' (2014) 32(Supp) *Social Cognition* 1, 6 (original emphasis); see also Dirk Wentura and Klaus Rothermund, 'Priming is not priming is not priming' (2014) 32 (Supp) *Social Cognition* 47, 54.
[138] Molden (2014), 6; citing Eliot R. Smith and Nyla R. Branscombe, 'Procedurally mediated social inference: The case of category accessibility effects' (1987) 23(5) *Journal of Experimental Social Psychology* 361;
[139] Wentura and Rothermund (2014), 54.

behaviours and the adoption and execution of goals – resulted in the next significant shift in theories of priming. Extrapolating from concurrent imaging evidence showing the activation of similar areas of the brain when imagining and performing the same actions,[140] "Direct Expression" theories have focused significantly on the first component of priming theories described above – the spreading of activation through semantically related networks. [141] In particular, these theories posit a direct link between representations that arise from stimuli that is perceived and representations of behaviour associated with that stimuli, such that the activation of the first automatically produces the activation of the second.

However, such direct expression theories are subject to the same limitations as their precursors concerning spreading activation. Without additional encoding subsequent to the initial priming effect, direct expression theories offer no explanation of longer priming effects for which the 'continued accessibility of the representation itself would appear to determine whether the associated behavior is enacted.'[142] Equally, a direct and automatic link between perception and behaviour without intermediate processing implies that primed behaviours should be expressed whenever activated, which is clearly contradicted by the significant interaction of a vast range of different moderating factors for virtually all priming effects. This requires explanation by way of some additional processing stage within which such moderators operate.[143]

Wheeler, DeMarree and Petty offer one theoretical explanation for this additional processing stage by reference to the "active-self", proposing that primes 'can increase the accessibility of primed and associated constructs, which in turn can shift the active self-

[140] For example, Wolfgang Prinz, 'Perception and action planning' (1997) 9(2) *European Journal of Cognitive Psychology* 129.
[141] For example, see Ap Dijksterhuis and John A. Bargh, 'The perception-behavior expressway: Automatic effects of social perception on social behavior' in Zanna M. P. (ed.), *Advances in Experimental Social Psychology: Vol. 33* (Academic Press 2001); John A. Bargh, Kay L. Schwader, Sarah E. Hailey, Rebecca L. Dyer and Erica J. Boothby, 'Automaticity in social-cognitive processes' (2012) 16(12) *Trends in Cognitive Sciences* 593.
[142] Molden (2014), 7.
[143] *Ibid*; for further criticism of direct expression theories, see Ben R. Newell and David R. Shanks, 'Prime numbers: Anchoring and its implications for theories of behavior priming' (2014) 32(Supp) *Social Cognition* 88.

concept.'[144] As the authors explain, this "Active-Self Theory" of priming parallels prior theories regarding chronic and temporary content within the active self-concept; chronic content being such characteristics of the self as goals, beliefs, values and self-knowledge which resides in long-term memory.[145] Crucially, the self-concept is dynamic; both chronic and temporary content is malleable and may change (albeit over different courses of time); what is more, different content may be activated at different times when making judgments or decisions, with chronic content being eponymously 'chronically available for activation.'[146] All other things being equal, the authors argue that content related to the self-concept that is most accessible is more likely to be applied to a judgment or decision; consequently, 'because priming affects the accessibility of information, it should be capable of affecting which information is in the active self-concept.'[147]

In support, Wheeler, DeMarree and Petty cite a broad range of research demonstrating where priming has been shown to affect subjects' active self-concept such as, for example,[148] where priming with images of standard or overweight people subsequently impacted upon subjects' own body image perception.[149] Furthermore, the active-self theory accounts for the wide range of priming moderators where previous direct expression theories fall down; the active self-account relates the various moderators to the effect that they have upon the active self-concept. For example, in relation to assimilation effects where behaviour is congruent with that implied by the prime, 'features that affect the extent to which primes can shift the active self-concept should

[144] S. Christian Wheeler, Kenneth G. DeMarree and Richard E. Petty, 'Understanding prime-to-behavior effects: Insights from the active-self account' (2014) 32(Supp) *Social Cognition* 109, 110; S. Christian Wheeler, Kenneth G. DeMarree and Richard E. Petty, 'Understanding the role of the self in prime-to-behavior effects: The active-self account' (2007) 11(3) *Personality and Social Psychology Review* 234; S. Christian Wheeler, Kenneth G. DeMarree and Richard E. Petty, 'The roles of the self in priming-to-behavior effects' in Tesser A., Wood J. V and Stapel D. A. (eds.), *On Building, Defending, and Regulating the Self: A Psychological Perspective* (Psychology Press 2005).

[145] For example, Hazel Markus and Ziva Kunda, 'Stability and malleability of the self-concept' (1986) 51(4) *Journal of Personality and Social Psychology* 858; Hazel Markus and Elissa Wurf, 'The dynamic self-concept: A social psychological perspective' (1987) 38(1) *Annual Review of Psychology* 299.

[146] Wheeler, DeMarree and Petty (2014), 111.

[147] *Ibid*.

[148] See further *Ibid*.

[149] Kerry Kawakami, Curtis E. Phills, Anthony G. Greenwald, Daniel Simard, Jeannette Pontiero, Amy Brnjas, Beenish Khan, Jennifer Mills and John F. Dovidio, 'In perfect harmony: Synchronizing the self to activated social categories' (2012) 102(3) *Journal of Personality and Social Psychology* 562.

likewise affect the magnitude of prime-to-behavior effects.'[150] Thus, for example,[151] it has been shown that people feel uncomfortable when they experience uncertainty regarding the self, which 'causes people to change their self-concepts in response to primes, depending on both the nature of the uncertainty and how the self is defined.'[152] Equally, in relation to contrast effects where behaviour is incongruent with that implied by the prime, the active self-account suggests that these result from primed constructs being activated which conflict with content in the self-concept.[153] For example, research demonstrates that both disliking[154] and feeling [155] distant from outgroups is more likely to promote contrast effects in response to outgroup primes.

Of greatest interest to the present thesis, the active-self theory accounts for how and why priming effects can operate outside of conscious awareness. Wheeler, DeMarree and Petty highlight research revealing an ironic property of the self-concept whereby 'implicit aspects of the self-concept are sometimes considered to reside outside of awareness[156] yet are highly accessible, such that they are automatically activated and can automatically influence responses.[157,158] This implies that it is not necessary for changes to the self-concept induced by priming to enter into conscious awareness in order to give rise to the effects demonstrated across prime-to-behaviour studies. Although, of course, the self-concept can be subject to conscious reflection and deliberative, controlled actions, the research shows that it is equally possible for the self-concept to be activated and guide actions implicitly. Thus, the 'self-related processes that direct behavior can

[150] Wheeler, DeMarree and Petty (2014), 113.
[151] See further *Ibid*.
[152] Kimberly Rios Morrison, Camille S. Johnson and S. Christian Wheeler, 'Not all selves feel the same uncertainty: Assimilation to primes among individualists and collectivists' (2012) 3(1) *Social Psychological and Personality Science* 118, 118.
[153] Wheeler, DeMarree and Petty (2014), 114.
[154] Cesario, Plaks and Higgins (2006).
[155] Alison Ledgerwood and Shelly Chaiken, 'Priming us and them: Automatic assimilation and contrast in group attitudes' (2007) 93(6) *Journal of Personality and Social Psychology* 940.
[156] Russell H. Fazio and Michael A. Olson, 'Implicit measures in social cognition research: Their meaning and use' (2003) 54(1) *Annual Review of Psychology* 297; Richard E. Petty, S. Christian Wheeler and Zakary L. Tormala, 'Persuasion and attitude change' in Millon T. and Lerner M. J. (eds.), *Handbook of Psychology: Volume 5 – Personality and Social Psychology* (John Wiley & Sons 2003).
[157] Thierry Devos, Que-Lam Huynh and Mahzarin R. Banaji, 'Implicit self and identity' in Leary M. R. and Tangney J. P. (eds.), *Handbook of Self and Identity* (2nd ed. The Guilford Press 2012); Anthony G. Greenwald and Shelly Farnham, 'Using the implicit association test to measure self-esteem and self-concept' (2000) 79(6) *Journal of Personality and Social Psychology* 1022.
[158] Wheeler, DeMarree and Petty (2007), 239.

simultaneously operate at varying levels of awareness and either in an automatic or controlled manner.'[159]

The active-self theory carries a number of implications that can be found expressed throughout social priming research. First, the vast range of moderators that have been identified clarify that 'behavioral priming effects are most likely to occur under specific conditions or among specific people.'[160] Second and relatedly, priming effects have almost certainly been overgeneralised in earlier research whereas, again, the discovery of a multitude of moderators reveals the subtlety and specificity of priming effects (and their moderators) to individuals and circumstances. And third, many factors have equally been revealed that can limit prime-to-behaviour effects; for example, when subjects lack self-confidence in their own thoughts priming is less likely to produce behavioural effects even though it continues to affect subjects' thoughts.[161]

An alternative "Resource Computational Theory" of priming automaticity is offered by Cesario and Jonas who propose that 'automatic social behaviors following priming [are] understood as the output of a computational process that assesses what a person can and cannot accomplish in response to others.'[162] This computation takes account of an individual's social resources – defining those possible behaviours that are likely to be successful with the support of others; bodily resources – defining possible behaviours that are likely to be successful in light of a person's bodily state and physiology; and structural resources – defining possible behaviours that are likely to be successful in view of the physical environment and availability of action-relevant objects.[163] These features are

[159] *Ibid*; citing Carolyn C. Morf and Walter Mischel, 'The self as a psycho-social dynamic processing system: Toward a converging science of selfhood' in Leary M. R. and Tangney J. P. (eds.), *Handbook of Self and Identity* (2nd ed. The Guilford Press 2012).
[160] Wheeler, DeMarree and Petty (2014), 115.
[161] Kenneth G. DeMarree, Chris Loersch, Pablo Briñol, Richard E. Petty, B. Keith Payne and Derek D. Rucker, 'From primed construct to motivated behavior: Validation processes in goal pursuit' (2012) 38(12) *Personality and Social Psychology Bulletin* 1659.
[162] Joseph Cesario and Kai J. Jonas, 'Replicability and models of priming: What a resource computation framework can tell us about expectations of replicability' (2014) 32(Supp) *Social Cognition* 124, 127.
[163] Tim W. Faber and Kai J. Jonas, 'Perception in a social context: Attention for response-functional means' (2013) 31(2) *Social Cognition* 301.

integrated under the resource computational model to select a given behavioural output in preference to other competing outputs.

The theory begins with the proposition that the perception of others – including those perceptions induced through priming – 'initiates self-regulatory systems to prepare the body for effective interactions with the target other.'[164] For effective interaction with others, it is necessary to incorporate information about the various aforementioned resources in order to calculate, prepare and execute the optimal (or preferred) behavioural option. Cesario and Jonas submit that this input of resource information is necessary in order for behaviour to be effectively regulated. They offer the example of both human and animal responses of aggression in contest situations which, in particular, takes account of information about the presence of social resources (*e.g.,* coalition members to support in contest), because such information changes both the likelihood of success and likely costs of aggressive behaviours.[165]

In support of resource computational theories of priming,[166] Cesario and Jonas focus on experiments utilising variables which should produce priming effects following this model, but would otherwise be regarded as irrelevant under other (predominantly direct expression spreading activation) models. For example, Cesario, Plaks and Higgins[167] produced a conceptual replication of earlier work by Bargh, Chen and Burrows[168] in which subjects were primed with racial stereotypes using images of African-American faces, discussed above in section 3.1.1 of this thesis. In Cesario, Plaks and Higgins' replication, however, subjects were specifically heterosexual men who were primed with the category of gay men, straight men, or no prime. With gay men almost universally being stereotyped with femininity and passivity, it was reasoned that a direct spreading

[164] Cesario and Jonas (2014), 127; citing Cesario, Plaks and Higgins (2006); Kai J. Jonas and Kai Sassenberg, 'Knowing how to react: Automatic response priming from social categories' (2006) 90(5) *Journal of Personality and Social Psychology* 709.

[165] Fessler and Holbrook (2013); Sarah Benson-Amram, Virginia K. Heinen, Sean L. Dryer and Kay E. Holekamp, 'Numerical assessment and individual call discrimination by wild spotted hyaenas, *Crocuta Mazarin*' (2011) 82(4) *Animal Behaviour* 743; Michael L. Wilson, Nicholas F. Britton and Nigel R. Franks, 'Chimpanzees and the mathematics of battle' (2002) 269(1496) *Proceedings of the Royal Society: Biological Sciences* 1107.

[166] See Cesario and Jonas (2014), 128 – 133.

[167] Cesario, Plaks and Higgins (2006).

[168] Bargh, Chen and Burrows (1996).

activation account of priming ought to result in decreased hostility. Conversely, if a self-regulatory response was being prepared following a resource computational model, it was expected that 'priming a negatively evaluated outgroup male… should result in more negative and aggressive responses', which was indeed the result of the study.[169]

Furthermore, again, the resource computation theory accounts for the automatic or nonconscious operation of prime-to-behaviour effects; indeed, the theory is presented as a model of automaticity. Cesario and Jonas conceive of the brain as a computational organ which receives informational input and then regulates the body in response according to sets of evolved psychological processes.[170] In order to do so effectively it must use information beyond solely that related to a target other; that is to say, 'stored knowledge about a social category member cannot be the sole determinant of behavioral output.'[171] This is why direct expression models of priming are insufficient; in the case of priming aggressive behaviour from racial stereotypes, the direct expression model would imply that the brain has evolved to respond aggressively whenever it perceives aggression, regardless of wider relevant factors such as the availability of one's coalition or access to defensive objects. Rather, under the resource computation model, 'a host of variables relevant to effective behavioral regulation *combine to determine automatic responses* to social category primes.'[172]

One of the most compelling theories of priming which draws from various aspects of those considered already is the "Situated Inference Model" from Loersch and Payne.[173] What is particularly powerful about this theory is that it can be used to account for the full range of perceptual, behavioural and goal priming effects that have been reviewed throughout this chapter of the thesis, as exemplified in the diagram below. The theory rests upon three premises: that priming increases the accessibility of certain related content or information which causes certain thoughts or emotions to be more likely to be

[169] Cesario and Jonas (2014), 128 – 129.
[170] *Ibid*., 128; citing Tooby and Cosmides (2005), (2016).
[171] *Ibid*.
[172] *Ibid*.
[173] Chris Loersch and B. Keith Payne, 'Situated inferences and the what, who, and where of priming' (2014) 32(Supp) *Social Cognition* 137; Chris Loersch and B. Keith Payne, 'The situated inference model: An integrative account of the effects of primes on perception, behavior, and motivation' (2011) 6(3) *Perspectives on Psychological Science* 234.

activated; that people typically assume that their thoughts and emotions concern whatever it is that they are attending to at that moment, even though they may be caused by something else; and that people tend to use those thoughts and emotions that are most accessible in order to guide their behavioural responses to different situations. Consequently, it is proposed that 'priming effects result when primes make certain ideas more likely to come to mind and those ideas are misattributed to one's own thoughts, interpreted in light of situational affordances.'[174]

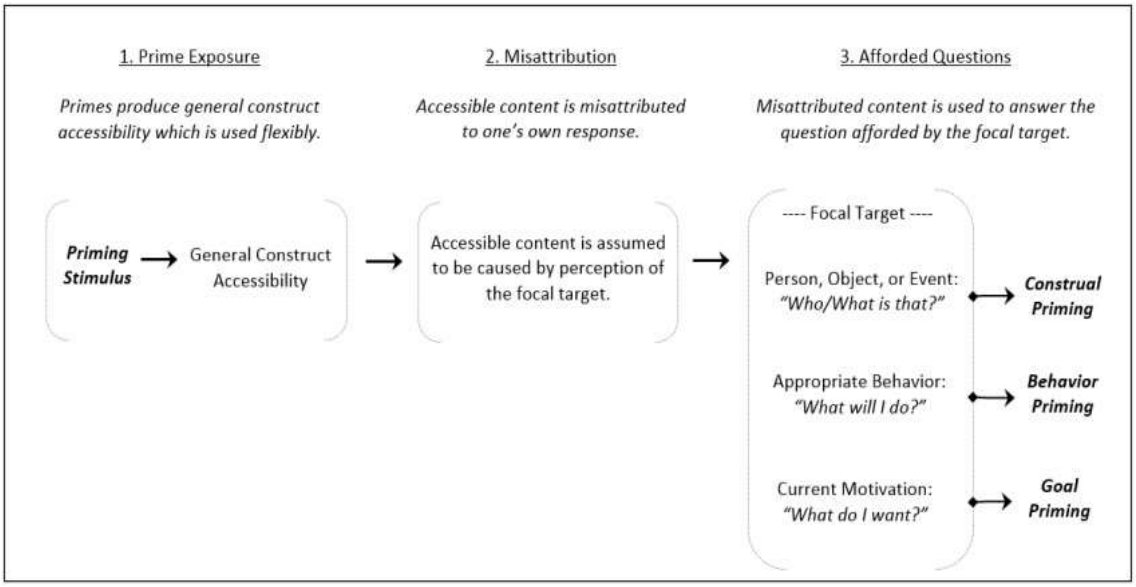


***Fig. g – Situated Inference Model of Priming.***[175]

Taking each proposition in turn, the situated inference model begins with ideas akin to the spreading activation models, whereby the priming stimulus activates mental content that is semantically, experientially or affectively related to that stimulus, rendering that content more accessible for judgment and decision-making. Thus, accessibility refers to the likelihood that certain mental content will be retrieved and applied to subsequent mental processes.[176] Crucially for the purposes of the present thesis, Loersch and Payne highlight that conscious awareness of the perception of priming stimuli or its activation

[174] Loersch and Payne (2014), 137.
[175] *Ibid*., 138.
[176] *Ibid*; citing Endel Tulving and Zena Pearlstone, 'Availability versus accessibility of information in memory for words' (1966) 5(4) *Journal of Verbal Learning and Verbal Behavior* 381; E. Tory Higgins, 'Knowledge activation: Accessibility, applicability, and salience' in Higgins E. T. and Kruglanski A. W. (eds.), *Social Psychology: Handbook of Basic Principles* (Guilford Publications 1996).

of mental content is not necessary; 'both conscious and nonconscious processing of a priming stimulus can produce the initial change in construct accessibility from which higher order social priming effects emerge.'[177]

Crucially, whereas direct expression accounts suggest that the increased accessibility of primed content leads automatically to behavioural outputs, the second stage of the situated inference model suggests that priming stimuli affect subsequent responses 'when this accessible content is mistakenly attributed to one's own internal thoughts and feelings about whatever is in the focus of attention.'[178] This follows because people typically act in ways that are consistent with their own thoughts and feelings regarding a given situation, judgment or decision (including when they have misattributed such thoughts or feelings as originating from themselves rather than some external stimulus). Loersch and Payne explain further, as people mistake information that has been activated through priming for being the result of their own thought processes, 'this mental content naturally becomes a source of bias in people's routine decision-making processes and is especially likely to be used to inform subsequent judgments, behavior, or motivation.'[179] Further, because accessible information is generally interpretated in relation to whatever is focal in attention, people are likely to regularly mistake the source of information activated by environmental stimuli.[180] Consequently, provided that priming stimuli is not so blatant or salient so as to be identified as the obvious source of its resulting mental content, information activated by such primes 'will be susceptible to the misattribution process.'[181]

---

[177] Loersch and Payne (2014), 138; citing Larry L. Jacoby, D. Stephen Lindsay and Jeffrey Toth, 'Unconscious influences revealed: Attention, awareness, and control' (1992) 47(6) *American Psychologist* 802; Anthony J. Marcel, 'Conscious and unconscious perception: Experiments on visual masking and word recognition' (1983) 15(2) *Cognitive Psychology* 197; Annette M. D. de Groot, 'The range of automatic spreading activation in word priming' (1983) 22(4) *Journal of Verbal Learning and Verbal Behavior* 417; Carol A. Fowler, George Wolford, Ronald Slade and Louis Tassinary, 'Lexical access with and without awareness' (1981) 110(3) *Journal of Experimental Psychology* 341.

[178] Loersch and Payne (2014), 139.

[179] Loersch and Payne (2011), 235.

[180] *Ibid*., citing; Gerald L. Clore and Karen Gasper, 'Feeling is believing: Some affective influences on belief' in Frijda N. H., Manstead A. S. R. and Bem S. (eds.), *Emotions and Beliefs: How Feelings Influence Thoughts* (Cambridge University Press 2000); E. Tory Higgins, 'The aboutness principle: A pervasive influence on human inference' (1998) 16(1) *Social Cognition* 173.

[181] *Ibid*., 237; citing Wendy J. Lombardi, E. Tory Higgins and John A. Bargh, 'The role of consciousness in priming effects on categorization: Assimilation versus contrast as a function of awareness of the priming task' (1987) 13(3) *Personality and Social Psychology Bulletin* 411; Norbert Schwarz and Gerald L. Clore,

The third premise under the situated inference model proposes that people will typically use the most accessible thoughts and emotions in order to guide behaviour; crucially, however, the meaning of any particular primed information to an individual will depend upon the particular situation within which they are using that information, *i.e.,* it depends on the particular question the individual is answering. For example, being asked to think about the personality traits of another person or oneself respectively affords questions such as "what type of person are they?" or "am I?". Loersch and Payne submit that, 'to the extent that prime-related content is misattributed to the focal target, these two situations will produce two distinct priming effects, differentially producing changes in other- versus self-perception.'[182] Whilst this example demonstrates the operation of the situated inference model in relation to perceptual (or construal) priming, the same model explains priming effects generally, including behavioural and goal (or motivation) priming.

Different situations (and, indeed, different experimental designs) give rise to different questions appropriate to the context – how to perceive something, how to behave and act, or what goal to adopt (*i.e.,* what one wants). It follows that mental content which becomes misattributed as an individual's own response to such questions will in turn affect the perceptual, behavioural or motivational inferences that person draws. To offer an example from behavioural priming, if mental content related to hostility is activated (*i.e.,* primed) and then misattributed as that person's own desire to aggress, behavioural priming effects such as administering more intense punishment to others may emerge.[183] The situated inference model therefore accounts for the broad diversity of priming effects which could emerge from the same priming stimulus according to its context; 'a single prime produces a myriad of downstream effects because its misattributed accessibility can have very different inferential implications across situations.'[184] Equally, the situated inference

'Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states' (1983) 45(3) *Journal of Personality and Social Psychology* 513.

[182] Loersch and Payne (2014), 139; citing Kenneth G. DeMarree and Chris Loersch, 'Who am I and who are you? Priming and the influence of self versus other focused attention' (2009) 45 *Journal of Experimental Social Psychology* 440.

[183] Loersch and Payne (2014), 139; citing Bargh, Lee-Chai, Barndollar, Gollwitzer and Trötschel (2001); Charles S. Carver, Ronald J. Ganellen, William J. Froming and William Chambers, 'Modeling: An analysis in terms of category accessibility' (1983) 19(5) *Journal of Experimental Social Psychology* 403.

[184] *Ibid*.

model accounts for the interaction of a broad range of moderators on priming effects, and Loersch and Payne explain a large body of experimental evidence concerning priming moderators within the theory.[185]

Once again – and most pertinent to the present thesis – the situated inference model significantly accounts for the nonconscious operation of priming effects. The processes described in spreading activation from a priming stimulus, misattribution of information to the self, and the application of accessible information to situational questions have all been shown to operate outside of conscious awareness.[186] Indeed, in a paragraph worth repeating, Loersch and Payne write:

> '[W]e view the basic process of using accessible information to infer the answer to environmentally afforded questions as a constant and obligatory aspect of the decision-making system, one that simply cannot operate at a solely conscious level. Because the environment continuously affords different questions as one seeks to understand the situation and determine how best to interact with the people and objects present, consciously attending to every decision would be untenable. Even without conscious involvement, however, the inference process we propose allows the mind to naturally integrate one's past learning history with the constraints of the current situation to guide behavior in a contextually appropriate manner. It is only because of the challenge of accurate source monitoring that this process introduces errors and produces priming effects.'[187]

[185] See *Ibid*., 140 – 145.

[186] *Ibid*., 139; citing Chris Loersch, Geoffrey R. O. Durso and Richard E. Petty, 'Vicissitudes of desire: A matching mechanism for subliminal persuasion' (2013) 4(5) *Social Psychological and Personality Science* 624; Chris Loersch and B. Keith Payne, 'On mental contamination: The Role of (mis)attribution in behavior priming' (2012) 30(2) *Social Cognition* 241; Christopher R. Jones, Russell H. Fazio and Michael A. Olson, 'Implicit misattribution as a mechanism underlying evaluative conditioning' (2009) 96(5) *Journal of Personality and Social Psychology* 933.

[187] Loersch and Payne (2014), 139 – 140.

### 3.1.4. *The Legal Relevance of Priming Research*

Priming effects are subtle but may be nonetheless powerful. Priming is subtle because the effects of any given priming stimulus are highly contextual, dependent upon a range of moderating factors according to both the idiosyncrasies of the individual, the environment, and the question or task to which that stimulus is being applied. Consequently, a priming stimulus may not necessarily even evoke a similar response from the same subject in different circumstances, and is less likely still to affect all subjects in a similar way. That notwithstanding, priming effects can be demonstrably powerful. Across the range of experiments considered in this chapter – from effects on perception and judgement, to effects on social impressions and responses, and the adoption and pursuit of goals – priming has been shown to cause affected individuals to reach perceptual, social and behavioural decisions which otherwise would have been different without the priming effects of certain stimuli.

This alone may not be particularly surprising; as a matter of survival, all animals and humans alike must make decisions *in response to* external stimuli and situations, whether searching for food, evading predation or interacting with others. That such external "priming" stimuli and environments should influence human decision-making is perfectly reasonable, if not expected. What is more surprising (and indeed contentious) is the manner in which priming stimuli appears to impact upon decision-making processes outside of conscious awareness. Such mental processes as decision-making, social interaction, and the selection, maintenance and pursuit of goals are traditionally assumed to require at least some minimal degree of conscious involvement. As the significant majority of priming experiments and their subsequent theoretical explanations reveal, however, these higher mental operations do indeed appear fully capable of operating outside of conscious awareness. This is not to say that consciousness has no role to play – a question that is explored more fully in relation to the *when* component of decision-making in chapter five, below. What is submitted, however, is that consciousness is not a *necessary* component of the decision-making process.

This latter point is particularly elaborated within the various different theories of priming. A common feature amongst each of the leading theories considered is that some or all of

the various stages are described as occurring automatically and outside of conscious awareness. In particular, the first stage of spreading activation – when the appearance of the priming stimulus causes the activation of relevant mental content which spreads amongst conceptually or semantically related content – is unanimously posited to occur unconsciously. However, subsequent stages within each theory are similarly largely explained through unconscious processes; whether suggesting that activated mental content is then computed into likely successful responses under the resource computational model, or positing that mental content triggered by a priming stimulus is then misattributed to the individual under the situated inference model, each of the priming theories substantially do not require conscious intervention in order to account for priming effects within decision-making processes.

What are the practical implications of the aforementioned research within the present thesis? The present chapter is concerned with the *what* component of a decision – *i.e.,* decisions about what to do in any given situation. Such decisions may be predominantly reactionary such as when deciding how to respond in a situation when faced with a number of options; equally, such decisions may be predominantly motivational whereby an individual selects, adopts and pursues a seemingly endogenous goal or objective. In this sense, the *what* component of a decision to commit some criminal action goes to the very heart of that criminal decision: the *what* component might be a decision to respond violently to an aggressor (a reactionary decision); or a decision to cheat on taxes (arguably a more motivational decision). Put differently, the *what* component of a given criminal decision will be that which typically contains the relevant subjective *mens rea* necessary for attributing legal responsibility for that criminal conduct – *i.e.,* the requisite intention, recklessness, dishonesty, or criminal knowledge, *etc.*

Firstly, the research from priming certainly suggests that occasions must arise where an individual's decision to commit a criminal act would likely not have arisen but for the influence of some priming stimuli. Naturally, any such effect would be dependent upon the appropriate concentration of circumstances and moderating factors which, for a particular individual exposed to particular stimuli and faced with a particular decision, facilitate that priming effect towards a criminal outcome. Recalling the multi-alternative

decision field theory of decision-making considered in section 2.3.1 of this thesis, above, different decision outcomes are represented in neuronal networks which competitively recruit evidence (increase activity) over time until reaching a threshold at which a decision outcome is reached. Priming can be understood as increasing the likelihood of a particular (potentially criminal) decision outcome in at least two ways. A prime might provide the initial stimulus to activate the representative neuronal network for a particular decision in the first place, effectively entering a particular decision option into the mix of those which are under consideration (and for which evidence is recruited) over time. Alternatively, or additionally, a priming stimulus may increase the very activity of a neuronal network representing a given decision, thereby acting as evidence in support of that decision outcome and rendering its ultimate selection more likely.

There is, however, an insurmountable practical limitation to this first conclusion (not least from a legal perspective), because it will be virtually impossible in any individual instance to prove that a decision or behaviour was the unescapable consequence of some exogenous priming influence. Simply, there exists insufficient objective access into the subjective decision-making processes to ever be able to demonstrate that a particular stimulus categorically resulted in a subsequent response outside of the conscious awareness and control of the individual concerned. In the example presented at the introduction to this thesis where a brain tumour demonstrably caused a defendant's criminal behaviour, significant proof of this causation was provided by the fact that the offensive behaviour disappeared with the excise of the tumour, reappeared when the tumour was found to be regrowing, and disappeared again following further surgery. Not only will it be intrinsically more difficult to identify the individual priming stimuli that have impacted upon any given decision, but proving the causative effect of such stimuli will be practically impossible. However, this should not be read as meaning that such priming stimuli – or, more accurately, the congregation of priming stimuli, their relevant moderators, and the circumstances of their operation on a particular decision – do not have a cumulative causative impact on the outcome of potentially criminal decisions.

The second and eminently more practical implication of the research in this chapter follows from the underlying theories of priming that have been considered, and the

implications for the role of consciousness in arriving at the *what* component of any given decision. In particular, each of the theories describes the automatic (or unconscious) operation of decision-making processes which culminate in the *what* component of a decision. Considering that the current conception of legal responsibility relies upon the actual subjective content of an individual's mind at the time they commit a criminal act, it follows from the theories presented in this chapter that people may be held responsible for decisions which were not only instigated from purely exogenous sources, but were further processed into bodily action entirely outside of conscious awareness and, therefore, the without realistic opportunity for conscious intervention or control.

Returning again to the clinical case presented at the introduction to this thesis, the patient's brain tumour was the demonstrable cause of significant changes in that individual's personality, behaviour and decision-making. In particular, the patient exhibited hyper-sexualised and inappropriate thoughts and desires which were grossly incommensurate with his ordinary character and previous history. So far as was discernible, the tumour not only caused (or, at the very least, greatly amplified) these thoughts and decisions, but also rendered him increasingly unable to exercise self-control and resist pursuing the sexual urges that he was experiencing. In this regard, there is a strong sense in which we intuit that the patient was not ultimately responsible for his deviant behaviour although, obviously, intervention through the criminal justice system became entirely warranted and necessary, not least for the protection of others.

A similar intuition arises in analogous cases where a person's reasoning, self-control and / or decision-making faculties are unduly influenced by some other medical condition or psychiatric disorder. For example, when an illness such as schizophrenia causes people to possess and pursue criminal intentions, the defence of insanity gives legal recognition to the fact that those criminal intentions have been caused by factors entirely outside of the individual's sphere of control or influence. In such examples, both physical and mental illness may not only influence the brain's decision-making mechanisms but hijack them entirely and, whilst intervention may again be warranted and necessary for the protection of that afflicted individual and others, both intuition and the law lead us not to hold such an individual responsible for their consequent actions.

Both the phenomenon of priming effects and, even more so, their underlying explanatory theories lead to the conclusion that an unknown quantity of our decisions may ultimately arise from exogenous causes over which we have little to no awareness and even less control. Moreover, all of our decisions result from mechanisms which similarly operate automatically and outside of conscious awareness. This is not to say that consciousness necessarily has no role to play, and this theme will continue to be expanded throughout the thesis. For present purposes, however, it may be said that the brain's decision-making apparatus can operate unconsciously and automatically, and does so for an unknown (but likely significant) proportion of the time. Chapters five and six of this thesis, concerning respectively the *when* and *whether* components of decision-making, further explore the timing and contribution of consciousness in decisions to act.

Returning to the *what* component, it flows from the implications above that an unknown (but likely significant) number of decisions result from processes over which we have little introspection or awareness. What is more, this follows not only for perceptual and reactionary decisions but also social and motivational decisions – including the formation of goals and intentions. If both intuition and the law teach that people are not responsible for decisions and actions arising from medical causes demonstrably outside of their subjective influence and control, why so should the same not follow for decisions and actions that are equally caused by other factors, (both exogenous and endogenous), which are outside of our control but so happen not to be medical in nature? Of course, the great *practical* difficulty with this perspective is that it is virtually impossible to distinguish or differentiate between those decisions which can reliably be attributed to the unconsciously processed influence of some exogenous cause (or prime), and those that can more fairly be attributed to the conscious, deliberative self and to which we more readily attach responsibility.

However, this perspective presents a more fundamental challenge to current conceptions of responsibility. The underlying theories of priming all propose that a significant component of the brain's decision-making apparatus can and does operate automatically and unconsciously, not least in relation to motivational decisions and the formation of goals. This means that an unknown proportion of intentions (criminal or otherwise) may

be the product of unconscious causes and processes over which the individual has no awareness and even less control, just as with the case of the man's deviant sexual behaviour resulting from a brain tumour or with a patient suffering from some mental illness which similarly disturbs their decision-making processes. The law's reliance upon subjective mental states such as intention, knowledge and recklessness (*i.e., mens rea*) as a determinative component of responsibility, therefore, becomes unreliable and somewhat arbitrary; an unknowable proportion of such criminal states of mind *will* be the result of unconsciously processed exogenous causes, which cannot readily be distinguished from endogenous, conscious and deliberate decisions.

Where the law abrogates responsibility when such inescapable influences over decision-making arise from identifiable medical causes, principle requires that responsibility should similarly be abrogated when decisions arise from the inescapable influence of any other cause over which the individual lacks awareness and / or control. It seems practically impossible to tell the difference with any reliability between volitional decisions resulting from endogenous motivations and conscious deliberation, from those triggered by exogenous causes and automatic processing. What is more, it may be the case that most or all decision-making in fact results from unconscious and automatic processes, allowing little room for intervention or control by the conscious self.

Thus, when the law imposes responsibility upon a person who commits a forbidden act intentionally or recklessly, there is no guarantee that such intentionality or recklessness can fairly be attributed to that individual, their endogenous motivations and conscious deliberation, or whether those mental states have arisen as the inescapable result of unconsciously processed exogenous causes for which the individual can scarcely be held responsible. Proof of subjective mental states as a requisite component of criminal liability therefore appears incapable of fulfilling the central purpose for which it is intended, *i.e.,* distinguishing between those who fairly ought or ought not to be held responsible for their decisions and actions.

## *3.2. Neural Correlates of Decision Outcomes Prior to Awareness*

An additional line of experimental research provides further evidence in support of the previous discussion and the unconscious, automatic operation of the mechanisms which result in the *what* component of a decision. This line of research flows from the paradigmatic work of Benjamin Libet in the 1970s and is discussed in greater detail in chapter five of this thesis, below, as it relates most pertinently to the *when* component of decision-making. Nevertheless, it is valuable to highlight the supportive role such evidence also plays in relation to the *what* component here discussed. In brief,[188] the original paradigm by Libet consists of subjects freely choosing when to perform a specific action such as flexing the wrist or pressing a button, whilst subjectively watching a makeshift clock and reporting the time at which they first experienced the subjective sensation of having reached a decision about when to act. Meanwhile, the exact time of physical movement was recorded by electromyogram ('EMG') whilst electroencephalogram ('EEG') recordings were taken from the subject's scalp.

In findings that have since been well replicated, Libet discovered unconscious brain activity which preceded and predicted not only the subject's subsequent physical movement but also their subjective experience of deciding to move. Whilst Libet's interpretation of the particular electrical signal that he recorded – the "readiness potential" ('RP') – is no longer accepted, more sophisticated replications of his paradigm utilising ever more precise brain imaging techniques have continued to reveal unconscious activity in the brain which both precedes and predicts decision outcomes across a range of different experimental settings and decision-making tasks.

Soon, Brass, Heinze and Haynes[189] conducted a broad replication of Libet's paradigm in 2008, the key differences being the use of fMRI as opposed to EEG in order to record

[188] See chapter five, below; Benjamin Libet, Elwood W. Wright Jr and Curtis A. Gleason, 'Preparation- or intention-to-act, in relation to pre-event potentials recorded at the vertex' (1983a) 56(4) *Electroencephalography and Clinical Neurophysiology* 367; Benjamin Libet, Elwood W. Wright Jr, Curtis A. Gleason and Dennis K. Pearl, 'Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act' (1983b) 106(3) *Brain: A Journal of Neurology* 623.

[189] Chun Siong Soon, Marcel Brass, Hans-Jochen Heinze and John-Dylan Haynes, 'Unconscious determinants of free decisions in the human brain' (2008) 11(5) *Nature Neuroscience* 543.

brain activity, and presenting subjects with a choice between left and right buttons to press. The experiment found two brain regions which 'encoded with high accuracy' whether the subject would choose the left or right button press by up to ten seconds ahead of the subject becoming consciously aware of making any such decision.[190] A related team consisting of Soon, He, Bode and Haynes[191] conducted a further conceptual replication using fMRI in 2013, wherein subjects had to choose at will between performing an addition or subtraction calculation on two numbers which were presented in each trial. Thus, the replication did not rely upon the subject preparing any motor action as the choice which they were asked to make was entirely abstract. Predictive activity was encoded in four brain regions similar to those found in the 2008 study, albeit the activity became reliable later from up to four seconds prior to the subjects' conscious awareness of reaching a decision.[192] These experiments provide further evidence for the unconscious operation of decision-making networks in the brain which appear significantly to be reaching a decisional outcome prior to subjects' conscious awareness of reaching any decision.

In 2011, Fried, Mukamel and Kreiman[193] had the opportunity to conduct a close replication of the Libet paradigm with twelve subjects implanted with depth electrodes in the brain for the treatment of intractable epilepsy, allowing for direct single cell measurements to be taken from discrete neuronal populations, and the most accurate possible measurement of the timing of brain activity so far as current technology allows. The results found that activity in small assemblies of single neurons could predict both the timing and direction of choice in the subjects' left or right button press. Moreover, preconscious activity was recorded from several-hundred milliseconds to several seconds prior to subjective awareness of having made a decision. An earlier 1991 experiment by Fried, Katz, McCarthy, Sass, Williamson, Spences and Spencer[194] is notable at this

[190] *Ibid*., 543 – 544.
[191] Chun Siong Soon, Anna Hanxi He, Stefan Bode and John-Dylan Haynes, 'Predicting free choices for abstract intentions' (2013) 110(15) *Proceedings of the National Academy of Sciences* 6217.
[192] *Ibid*., 6218 – 6219.
[193] Itzhak Fried, Roy Mukamel and Gabriel Kreiman, 'Internally generated preactivation of single neurons in human medial frontal cortex predicts volition' (2011) 69(3) *Neuron* 548.
[194] Itzhak Fried, Amiram Katz, Gregory McCarthy, Kimberlee J. Sass, Peter Williamson, Susan S. Spencer and Dennis D. Spencer, 'Functional organization of human supplementary motor cortex studied by electrical stimulation' (1991) 11(11) *The Journal of Neuroscience* 3656.

juncture, in which subjects also included clinical patients undergoing brain surgery for intractable epilepsy. In this experiment, electrical stimulation was applied directly to various parts of the brain, most of which elicited overt motor movements from the subjects or, less frequently, the subjective experience of moving without any corresponding overt movement, or experiences of urging and anticipating movement.[195]

From each experiment, Fried *et. al.* concluded that the experience of will or intention to move itself 'emerges as the culmination of premotor activity… starting several hundreds of [milliseconds] before awareness.'[196] In addition, whilst these experiments provide evidence for the unconscious processing of decisions prior to conscious awareness of reaching a choice, they further begin to explain the very experience of subjective intention as being an inherent part of the premotor activity which culminates in action. In this sense, a subjective state of intention does not necessarily initiate or cause action *ab initio*, but is one of the stages in the process of enacting an unconsciously reached decision to perform an action, the initiation of which occurs earlier and prior to conscious awareness.

An illuminating set of experiments by Wisniewski, Goschke and Haynes,[197] Zhang, Hughes and Rowe,[198] and Zhang, Kriegeskorte, Carlin and Rowe[199] each used fMRI to explore the brain regions involved in forming intentions and, in particular with regards to Wisniewski *et. al.*, the differences between endogenously generated and externally cued intentions. Each experiment revealed that the same brain network was engaged in intentional choices, regardless of whether they were freely endogenously generated or cued by external stimuli. These findings are particularly illuminating in light of the research surrounding priming effects which, for example, similarly reveal that motivational decisions – the adoption and pursuit of goals – operate in the same manner whether endogenously adopted or exogenously primed. Again, this suggests that

[195] *Ibid*., 3658.
[196] Fried *et. al.* (2011), 557.
[197] David Wisniewski, Thomas Goschke and John-Dylan Haynes, 'Similar coding of freely chosen and externally cued intentions in a fronto-parietal network' (2016) 134 *NeuroImage* 450.
[198] Jiaxiang Zhang, Laura E. Hughes and James B. Rowe, 'Selection and inhibition mechanisms for human voluntary action decisions' (2012) 63(1) *NeuroImage* 392.
[199] Jiaxiang Zhang, Nikolaus Kriegeskorte, Johan D. Carlin and James B. Rowe, 'Choosing the rules: Distinct and overlapping frontoparietal representations of task rules for perceptual decisions' (2013) 33(29) *Journal of Neuroscience* 11852.

intentionality itself represents a relevant stage within the decision-making process, but does not reflect any distinction between those intentions that are produced freely and endogenously and those which are externally cued or primed.

Finally, Lau and Passingham[200] present an fMRI study in which subjects were instructed to perform *either* a phonological judgement or a semantic judgment in relation to words presented in each trial. In some trials, however, subjects were primed to prepare the opposite task to that which they were explicitly instructed to complete, the effect of which was to subsequently impair their performance on the instructed task. When subjects had been so primed, the fMRI results revealed a decrease in neural activity in areas of the brain relevant to the instructed task and a corresponding increase in activity in those areas related to the primed task. This demonstrated that the subjects' brains had actually been engaged on the wrong (primed) task, and that subjects were not simply being distracted by the priming stimulus. These findings provide yet further evidence for the way in which the *what* component of the decision-making processes can be influenced by stimuli processed outside of conscious awareness, again revealing the automatic and unconscious operation of the associated decision-making networks in the brain.

## 3.3. From Priming and Predicting Decisions to Legal Responsibility

The *what* component of a decision contains the very essence of a potentially criminal decision to act. That is to say, a decision to fight (in attack or defence) or to flee is a decision about *what* to do; similarly, a decision to steal something, to lie on a tax return or insurance claim, to touch another person sexually, and to drive after drinking or over the speed limit, are all decisions about *what* to do in certain situations. Such decisions seemingly can and do result from a conscious deliberative process, such that an individual can consciously, "willingly" intend to do these and other criminal acts – consciousness is not, as of yet, excluded entirely from decision-making processes. What this chapter demonstrates, however, is that the *what* component of a decision may equally arise

[200] Hakwan C. Lau and Richard E. Passingham, 'Unconscious activation of the cognitive control system in the human prefrontal cortex' (2007) 27(21) *Journal of Neuroscience* 5805.

entirely unconsciously and / or automatically. Where the law asks whether a person *intended* to do a criminal act, or did so *knowingly* or being *aware of risks etc.,* this investigation does not necessarily distinguish between those who have formed a *what* decision consciously and intentionally in the manner that the law presumes, and those whose decision about *what* to do arose unconsciously and / or automatically.

The legal concept of *mens rea* is one of the (usual) requisite components for criminal responsibility, and requires that the defendant possessed a particular "guilty" state of mind when they committed a given offence, such as an *intention* to injure another, *recklessness* as to harm flowing from their actions, or *knowledge* that a criminal state of affairs exists, *etc*. It is generally assumed that if people consciously exhibited such criminal states of mind at the time of their offending, then it existed within their sphere of control to choose not to commit a given criminal act, as people are presumed to possess free will and volitional control.

However, if intention, recklessness, or even knowledge can be activated, processed and determinatively lead to particular actions and behaviours entirely unconsciously, this presumption is undermined. Similarly, it is frequently reasoned that subjective states of mind denote guilt because they reflect the outcome of a person's deliberative choice which, knowing the illegal consequences of their actions, they could have concluded differently. However, again, the opportunity for such criminal states of mind to arise as the direct result of exogenous cues (or primes) being processed unconsciously undermines the argument that intention, recklessness and other subjective mental states necessarily reflect the outcome of some conscious and deliberative process.

Furthermore, the latter studies under consideration revealed similar brain networks involved in reflecting intentions, whether they arise endogenously or are externally cued. Indeed, the evidence suggest that the sensation of intention itself is a component part of the route from an initial decision being taken (likely unconsciously) to its performance through motor actions. In this case, however, the legal focus on intention and other such subjective states of mind cannot necessarily tell us anything about whether a person has acted on their own motives, as a result of a conscious and deliberative process, or acted

according to some externally cued and unconsciously processed stimulus. Interpreting such evidence at its high point, many scholars conclude that the entire decision-making apparatus operates without the *necessary* requirement for any conscious involvement. And, indeed, the proceeding chapters of this thesis continue to describe an ever-smaller role that consciousness might *necessarily* play in decision-making processes, notwithstanding that consciousness may still improve the performance of otherwise unconsciously operating processes.

What can fairly be concluded in this present chapter is that an unknown proportion of the *what* component of our decisions – *i.e.,* our intentions to do *x* or *y* – arise from external cues and unconscious processes over which we have little to no subjective insight or conscious control. Consequently, the legal focus on such subjective states of mind as intentionality, recklessness, knowledge and dishonesty, *etc.* likely tells us little about whether or not a person can truly be said to be consciously and deliberatively responsible for their decisions and actions. Proof of such subjective states of mind reveals nothing of whether they are a reflection of an individual's personal and deliberate choice, or the happenstance result of automatic mechanisms processing external cues and stimuli.

# 4. The *How* Component and Sense of Agency

> 'This is the excellent foppery of the world, that when we are sick in fortune – often the surfeits of our own behaviour – we make guilty of our disaster the sun, the moon and stars, as if we were villains by necessity, fools by heavenly compulsion, knaves, thieves and treachers by spherical predominance, drunkards, liars and adulterers by an enforced obedience of planetary influence, and all that we are evil in, by a divine thrusting on: an admirable evasion of whoremaster man, to lay his goatish disposition on the charge of a star!'
>
> - William Shakespeare, 1606.[1]

The question of *what* to do will often, if not almost always, be inextricably linked to the related question of *how* to do the thing that is chosen; if a certain option (*i.e., what*) is easy or difficult (*i.e., how*) relative to other options, that will be an important consideration in the ultimate decision of which option to choose. Indeed, the first section of this chapter reveals the close connection between the *what* and *how* components in the brain, proposing that the brain in fact prepares multiple plans for different actions (*how*) as part of the process of deciding *what* to do in a given situation. This may be relevant to questions of responsibility for action because, it follows that the brain may unconsciously prepare the actions which would result in the commission of a criminal offence, simply as part of the very process of deciding what to do. In relevant circumstances, this would place individuals in a state of readiness to potentially commit some criminal offence, rendering it more likely for that prepared action to actually be initiated, and again diminishing the opportunity for active, conscious intervention by the individual.

The question of *how* to do a particular act is inherently concerned with action planning in the brain. Of particular relevance to the topic of responsibility, the second section of this

[1] William Shakespeare, *King Lear* (Bate J. and Rasmussen E. (eds.), Macmillan Publishers 2009), 41.

chapter reveals how the human sense of agency likely emerges as a function of action planning. More specifically, it is likely that the sense of agency arises when performed actions / outcomes conform with their original plans, whereas feelings of agency diminish when bodily actions do not conform with plans as intended, for example, when somebody accidentally knocks over a glass they were otherwise intending to grasp and pick up. From this perspective, the subjective sensation of agency or responsibility for personal action does not necessarily provide any marker that those actions were consciously and deliberately chosen by an individual. Whereas sense of agency may indicate that a person has performed some bodily motion properly and in accordance with a relevant plan of action in the brain, agency does not indicate that that plan of action was consciously selected and approved by the individual. Instead, following from the previous chapter of this thesis, such action plans may be triggered, developed and initiated into physical action outside of conscious awareness or control, yet may still result in a sense of agency or ownership over the resultant actions provided that they conform with their underlying action plans in the brain.

## *4.1. The Connection between "What" and "How" Components*

Discussed in section 2.2.1 of this thesis, above, experimental research from Ariani, Wurm and Lingnau [2] disassociated areas across the fronto-median wall which are either commonly or distinctly associated with internally and externally triggered *plans* to make particular actions – *i.e.,* the *how* component of a decision. In an fMRI study, subjects were either instructed to perform, or could choose freely between, one of three different hand actions, and were separately instructed when to plan the performance of that action and when to carry it out, thus controlling the *what, when* and *whether* of each decision and isolating a period of time for the *how* component. Further, by also investigating free choice trials where the subjects selected the action themselves, it was possible to compare and contrast areas involved in planning endogenously and exogenously selected actions.

[2] Giacomo Ariani, Mortiz F. Wurum and Angelika Lingnau, 'Decoding internally and externally driven movement plans' (2015) 35(42) *Journal of Neuroscience* 14160.

Three key results were obtained: first, activity in the superior parietal lobule ('SPL'), intraparietal sulcus ('IPS'), dorsal premotor cortex ('PMd') and primary motor cortex ('M1') contralateral to the hand being moved was found to be associated with action planning regardless of whether that action was internally or externally cued. Second, activity was found in the contralateral ventral premotor cortex ('PMv'), dorsolateral prefrontal cortex ('dlPFC'), supramarginal gyrus ('SMG'), ipsilateral posterior intraparietal sulcus ('IPS'), posterior superior temporal gyrus ('STG'), and posterior middle temporal gyrus ('MTG') for internally-, but not externally-, driven movement planning. Third, activity was recorded in the bilateral sensory motor area ('SMA') and pre-SMA for encoding externally-driven movement plans.[3]

However, a number of findings from this and surrounding researching highlights notable overlap with certain brain regions which are also engaged with the *what* component of a decision, suggesting the possibility for some direct, functional connection between the *what* and *how* components of decision-making. For example, Ariani, Wurm and Lingnau decoded activity in the pre-SMA and SMA for planning instructed movements, in agreement with previous finings.[4] The fact that the same areas are specifically not engaged in the planning of freely chosen movements would indeed suggest that these areas are more discretely concerned with action planning. Meanwhile, in contrast, prior research has similarly linked activity in the SMA to the voluntary selection of actions and self-initiated movements,[5] which would comprise the *what* component of a decision.

[3] *Ibid*., 14168.

[4] Jason P. Gallivan, D. Adam McLean, J. Randall Flanagan and Jody C. Culham, 'Where one hand meets the other: limb-specific and action-dependent movement plans decoded from preparatory signals in single human frontoparietal brain areas' (2013) 33(5) *Journal of Neuroscience* 1991; Egbert Hartstra, Florian Waszak and Marcel Brass, 'The implementation of verbal instructions: Dissociating motor preparation from the formation of stimulus-response associations' (2012) 63(3) *NeuroImage* 1143; Jason P. Gallivan, D. Adam McLean, Kenneth F. Valyear, Charles E. Pettypiece and Jody C. Culham, 'Decoding action intentions from preparatory brain activity in human parieto-frontal networks' (2011) 31(26) *Journal of Neuroscience* 9599.

[5] For example, Jiaxiang Zhang, Nikolaus Kriegeskorte, Johan D. Carlin and James B. Rowe, 'Choosing the rules: Distinct and overlapping frontoparietal representations of task rules for perceptual decisions' (2013) 33(29) *Journal of Neuroscience* 11852; Jiaxiang Zhang, Laura E. Hughes and James B. Rowe, 'Selection and inhibition mechanisms for human voluntary action decisions' (2012) 63(1) *NeuroImage* 392; Itzhak Fried, Roy Mukamel and Gabriel Kreiman, 'Internally generated preactivation of single neurons in human medial frontal cortex predicts volition' (2011) 69(3) *Neuron* 548; Hakwan C. Lau, Robert D. Rogers, Narender

Also discussed above in section 2.3.2 of the present thesis, Cisek offers a computational model of decision-making which proposes a particularly close relationship between the *what* and *how* components of a decision.[6] He begins by noting that an animal must solve the related problems of deciding what to do and how to do it in order to achieve a behavioural goal, and proposes that this takes place through an interconnected mechanism within which 'a cell tuned to a specific value of some spatial parameter of movement is active in proportion to sensory and cognitive information favoring the selection of actions with that specific parameter value.'[7] More simply, the model proposes that the selection of *what* to do and *how* to do it is processed in parallel, with the variables for each of the available *what* and *how* options co-interacting. More simply still, the brain considers *how* to achieve different options as part of deciding *what* to do and, concurrently, takes account of the relative value of options for *what* to do when considering *how* to achieve each option. Section 2.3.2 of this thesis proposes how Cisek's theory might be extended to each of the *what, how, when, whether* and *why* components of a decision to account for the parallel and interconnected computation of each component in the brain as an ultimate rounded decision is reached.

Remaining with Cisek's theory connecting the *what* and *how* components in particular, a series of experimental research by Gallivan *et. al.*[8] provides empirical evidence for Cisek's thesis in action. In particular, the researchers were investigating the question of whether, when selecting between multiple available actions, the brain visually encodes

---

Ramnani and Richard E. Passingham, 'Willed action and attention to the selection of action' (2004b) 21(4) *NeuroImage* 1407.

[6] Paul Cisek, 'Integrated neural processes for defining potential actions and deciding between them: A computational model' (2006) 26(38) *Journal of Neuroscience* 9761; Paul Cisek, 'Cortical mechanisms of action selection: The affordance competition hypothesis' (2007) 362(1485) *Philosophical Transactions of the Royal Society: Biological Sciences* 1585; Paul Cisek and John F. Kalaska, 'Neural mechanisms for interacting with a world full of action choices' (2010) 33(1) *Annual Review of Neuroscience* 269.

[7] Cisek (2006), 9761.

[8] Jason P. Gallivan, Brandie M. Stewart, Lee A. Baugh, Daniel M. Wolpert and J. Randal Flanagan, 'Rapid automatic motor encoding of competing reach options' (2017) 18(7) *Cell Reports* 1619; Jason P. Gallivan, Natasha A. R. Bowman, Craig S. Chapman, Daniel M. Wolpert and J. Randall Flanagan, 'The sequential encoding of competing action goals involves dynamic restructuring of motor plans in working memory' (2016) 115(6) *Journal of Neurophysiology* 3113; Jason P. Gallivan, Kathryn S. Barton, Craig S. Chapman, Daniel M. Wolpert and J. Randall Flanagan, 'Action plan co-optimization reveals the parallel encoding of competing reach movements' (2015) 6 *Nature Communications* 7428; Brandie M. Stewart, Jason P. Gallivan, Lee A. Baugh and J. Randall Flanagan, 'Motor, not visual, encoding of potential reach targets' (2014) 24(19) *Current Biology* R953.

the location of action targets in order to guide movement theretoward, or whether the brain actually prepares motor representations for each possible action. Subjects performed a variation of the "go-before-you-know" paradigm in which they were required to initiate a movement towards two or more possible targets *before* knowing which was the actual target. Previous experiments have shown that people will generally aim for a midpoint between the distribution of potential targets before correcting for the actual target,[9] which reduces the overall cost of corrective actions needed later.[10] Using a joystick, subjects were instructed to direct a virtual cursor towards targets on a screen at -30°, 0° and +30°, either knowing the target beforehand (single target trials) or towards two or more targets but without knowing the actual target until after they had initiated movement (go-before-you-know trials). Next, the experimenters gradually adapted the subjects to visuomotor rotations with the effect that they would unwittingly produce identical straight-forward movements for targets situated at both 0° and +30°.

From here, as Gallivan *et. al.* explain, the 'visual and motor encoding hypotheses now make different predictions with respect to the initial movement direction in two-target trials.'[11] Following the visual encoding hypothesis, the subjects should continue to motion straight forward (*i.e.,* 0°) on go-before-you-know trials because this remains the *visually* averaged direction and reaches towards 0° were unaffected by the visuomotor rotation adaptations. Conversely, following the motor encoding hypothesis, the initial movement would be influenced by the adaptation and subjects should reach towards the midway point between the un-adapted target at -30° and the target at +30° which has in fact been adapted to 0° - *i.e.,* they would reach for -15°. The results showed that, after subjects were adapted to the visuomotor rotations, their reaches on go-before-you-know trials were shifted significantly to the left, consistent with the motor encoding hypothesis

[9] Jason P. Gallivan, Craig S. Chapman, Daniel K. Wood, Jennifer L. Milne, Daniel Ansari, Jody C. Culham and Melvyn A. Goodale, 'One to four, and nothing more: Nonconscious parallel individuation of objects during action planning' (2011) 22(6) *Psychological Science* 803; Craig S. Chapman, Jason P. Gallivan, Daniel K. Wood, Jennifer L. Milne, Jody C. Culham and Melvyn A. Goodale, 'Short-term motor plasticity revealed in a visuomotor decision-making task' (2010) 214(1) *Behavioural Brain Research* 120.

[10] Adrian M. Haith, David M. Huberdeau and John W. Krakauer, 'Hedging your bets: Intermediate movements as optimal behavior in the context of an incomplete decision' (2015) 11(3) *PLOS Computational Biology* e1004171; Brandie M. Stewart, Lee A. Baugh, Jason P. Gallivan and J. Randall Flanagan, 'Simultaneous encoding of the direction and orientation of potential targets during reach planning: Evidence of multiple competing reach plans' (2013) 110(4) *Journal of Neurophysiology* 807.

[11] Gallivan *et. al.* (2017), 1620.

and Cisek's broader thesis that, '*prior to target selection and subsequent movement execution*, competing potential reach targets are rapidly and automatically transformed into corresponding motor representations.'[12]

Gallivan *et. al.* add further that their behavioural findings offer a 'strong interpretation' of recent neurophysiological studies which suggest that multiple spatial goals are represented in the sensorimotor areas of the brain,[13] 'namely that this activity directly reflects movement-related parameters associated with these goals.' [14] Finally, contemporary research from Cos, Medleg and Cisek[15] explores the biomechanics of how people move their arms in pursuit of a free choice of targets. Beginning with prior neurophysiological findings suggesting that sensorimotor areas of the brain simultaneously represent different courses of action,[16] and considering Cisek's theory that these representations compete to reach a final decision and overt execution,[17] Cos, Medleg and Cisek add:

> 'As decision-making variables are computed and gradually fine tuned, they can bias the competition in favor of the "better" choice. Our data suggest that this bias includes information about the biomechanical properties of the motor apparatus as well as about the difficulty of controlling a given

[12] *Ibid*., 1623.
[13] Tineke Grent-'t-Jong, Robert Oostenveld, W. Pieter Medendorp and Peter Praamstra, 'Separating visual and motor components of motor cortex activation for multiple reach targets: A visuomotor adaptation study' (2015) 35(45) *Journal of Neuroscience* 15135; Christian Klaes, Stephanie Westendorff, Shubhodeep Chakrabarti and Alexander Gail, 'Choosing goals, not rules: Deciding among rule-based actions plans' (2011) 70(3) *Neuron* 536.
[14] Gallivan *et. al.* (2017), 1623.
[15] Ignasi Cos, Farid Medleg and Paul Cisek, 'The modulatory influence of end-point controllability on decisions between actions' (2012) 108(6) *Journal of Neurophysiology* 1764.
[16] Markus A. Baumann, Marie-Christine Fluet and Hansjörg Scherberger, 'Context-specific grasp movement representation in the macaque anterior intraparietal area' (2009) 29(20) *Journal of Neuroscience* 6436; Paul Cisek and John F. Kalaska, 'Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action' (2005) 45 (5) *Neuron* 801; Paul Cisek and John F. Kalaska, 'Simultaneous encoding of multiple potential reach directions in dorsal premotor cortex' (2002) 87(2) *Journal of Neurophysiology* 1149; Robert M. McPeek and Edward L. Keller, 'Superior colliculus activity related to concurrent processing of saccade goals in a visual search task' (2002) 87(4) *Journal of Neurophysiology* 1805.
[17] Cisek and Kalaska (2010); Cisek (2007); Alexandre Pastor-Bernier and Paul Cisek, 'Neural correlates of biased competition in premotor cortex' (2011) 31(19) *Journal of Neuroscience* 7083.

movement and that both of these are at least partially estimated before movement onset.'[18]

The above research may be considered together: there is evidence for a degree of neuroanatomical overlap between the regions of the brain engaged with deciding *what* to do and preparing the plans for *how* to do it. Further, there is behavioural evidence which supports the suggestion that the brain prepares multiple potential movements (*how*) reflecting different possible choices (*what*) as part of the overall process of deciding what to do. That is to say two things: first, the relative ease and associated costs of each potential action plan are factors computed as part of a decision of what to do; and second, during the process of deciding what to do, the brain is already preparing the various motor actions necessary to execute any number of the potential decision outcomes under consideration. Finally, each of these findings can be explained within Cisek's distributed consensus model of decision-making, whereunder different goals (*what*) and their corresponding actions (*how*) are represented across many levels, and decisions are reached through 'competition at multiple levels of representation' simultaneously computing (at least) both the *when* and *how* components of any decision.[19]

To some degree there is an inherent logic behind these findings; in deciding what to do at any given moment, both *how* any of the various options might be achieved – (if, indeed, they *can* be achieved at all) – and the relative ease or costs associated with each option are eminently relevant considerations in the ultimate decision of *what* to do. To exemplify, suppose one is deciding what drink to get – water from the tap, a coffee from the machine, or a beer from the bar. If the coffee machine is broken, the *how* of obtaining coffee becomes impossible and this option should be removed from consideration; equally, if the bar is closed, the *how* of obtaining beer becomes impossible and this option should similarly be removed. In this manner, considering how any given option might be achieved and the relative costs associated therewith becomes a crucial factor to take into account when deciding which option to choose in the first place. The evidence suggests

[18] Cos, Medleg and Cisek (2012), 1778.
[19] Paul Cisek, 'Making decisions through a distributed consensus' (2012) 22(6) *Current Opinion in Neurobiology* 927, 928; see further section 2.3.2 of this thesis, above.

that the brain has evolved according to this logical truism such that *how* to achieve any particular goal becomes computed as part of the very process of deciding *what* goal to pursue.

### 4.1.1. *The Legal Relevance of Connections between the "What" and "How" Components*

As previously considered, planning to do something can encompass both the long-term deliberation of the various steps involved in fulfilling a distant goal, and the unconscious preparation of the immediate motor actions necessary to achieve the next step towards that particular goal. For the purposes of the law, however, it is the latter conception of planning which is of arguably greater importance. Many, if not most, people fantasise or imagine plans at some point in their life which would be criminal if actually put into action, from a child who thinks about the relatively mundane act of shoplifting some sweets, to an adult who imagines exacting their revenge against a hated colleague. Such plans may become even more detailed and elaborate within the arts, for example, where writers and directors of crime fiction devise and plan all manner of criminal acts.

Of course, one critical feature of such aforementioned "planning" is that it falls within the realm of fantasy – the examples of planning given above are not generally accompanied by any genuine intention to put those plans into action and commit a criminal act. Even more crucially for the purposes of the law, any such examples of deliberate planning over time do not become criminal unless and until some minimal action is initiated towards their completion, whether this consists of sharing those plans with another for the purposes of criminal conspiracy, gathering equipment or materials such as for the offence of preparing terrorist activities, or simply saying something threatening or abusive to another for the offences of assault and hate speech respectively. The point being that planning alone, whether deliberated consciously over time or unconsciously in the moment of initiating action, cannot form the basis of criminal responsibility without some further action, however minimal. This in fact renders the *whether* component of decision-making arguably *the* most critical to the question of responsibility, insofar as the decision to actually execute any plan must always be the

final step before a person has done something for which they are to be held criminally responsible.

There is a further implication which may also be drawn from the above discussion. Just as the *whether* component of a decision is the final boundary which must be crossed in order to result in potentially criminal *behaviour* for which an individual may be held responsible, so it is the latter aspects of the *how* component which are similarly more important to the question of responsibility. That is to say, it is the motor actions which the brain is preparing (*i.e.,* the *how* component) immediately before the *whether* boundary is crossed which will ultimately determine the outcome of our subsequent actions. This is important to note when one considers the proportion of criminal offences which are not carried out as the result of careful planning and consideration, but which are instead opportunistic, reactionary, situational or occur "in the heat of the moment". It is likely not possible in practice to quantify the precise proportion of criminal offences that occur as a result of deliberate planning rather than opportunism or purely situational factors, but certainly the latter is unlikely to be rare. Indeed, some scholars of criminal theory propose that "opportunity" and opportunism 'plays a role in causing all crime.'[20]

The discussion in section 4.1, above, essentially proposes that, when engaged in the process of deciding what action to perform in a given moment, the brain actually prepares the various motor actions associated with each of the options for action under consideration, and factors in the expected costs associated with each action plan as part of the process of deciding which plan to ultimately pursue and initiate through physical action. What is more, as has already been suggested and is elucidated further in section 4.2 of the present chapter, below, this preparation of different competing motor actions as part of reaching a decision in the moment occurs outside of conscious awareness.

The above can be exemplified in a number of hypothetical but realistic scenarios. Consider the man in a bar who is knocked, causing him to spill his drink over himself. In the moments as he turns to face his *potential assailant*, it is proposed that the brain would

[20] Marcus Felson and Ronald V. Clarke, 'Opportunity makes the thief: Practical theory for crime prevention' (Home Office Policing and Reducing Crime Unit, Police research series paper 98, 1998), *v*.

be preparing *both* the actions to engage in discourse or to aggressively strike back, not yet knowing whether he is under attack. On the one hand, the man needs a few milliseconds more to face the person who knocked him and assess whether this was an accident or if he is under some sustained attack. On the other hand, the man needs to be able to respond quickly in the circumstances that he is under attack. The brain seemingly addresses this need for rapid reaction by preparing the appropriate motor responses for both outcomes – *i.e.,* the brain simultaneously prepares for both options of engaging in discourse and responding with violence.

From this perspective it is easy to appreciate how, in the circumstances described (and not least with the involvement of alcohol), what begins as an unintended accident can easily descend into violence amongst even ordinarily peaceful and law-abiding individuals. With the brain having prepared both the non-confrontational and confrontational responses, it perhaps only requires one further mistake, misapprehension or provocation for a ready and waiting violent response to be initiated. This is all the more illuminating in light of research considered in sections 3.1 and 3.1.1 revealing how people form automatic and unconscious impressions of others and their goals, and how these impressions can themselves be influenced by other cues or stimuli acting as a prime. Thus, the argument follows, the person knocked in a bar may not only be unconsciously preparing to respond to aggression in kind, but may also reach their assessment of the "aggressor" in a largely automatic and unconscious manner, and perhaps under the influence of other exogenous priming stimuli. Consequently, it is possible to appreciate how even an ordinarily law-abiding person could find themselves responding violently to a perceived aggressor, based largely upon the automatic and unconscious processing of their perception of that aggressor and their interpretation of the aggressor's act of knocking into them.

Consider instead the woman driving from work as she turns onto a clearly empty stretch of road, tired and eager to return home; she could maintain her steady course or, upon seeing the empty road ahead, might press the accelerator harder and break the speed limit. Again, it is argued that the brain could be preparing both of these actions – *i.e.,* to maintain or increase pressure on the accelerator. Our hypothetical woman might again be an

ordinarily law-abiding citizen; but, already in a state of preparation to increase the pressure with her foot, it is readily appreciable how perhaps a flush of eagerness or a mere lapse in self-control could operate as the final trigger to initiate her criminal act of breaking the speed limit. Indeed, this is likely a scenario which may be all too familiar to a great majority of drivers at some point in their lives.

As a final example, consider a waitress in debt and struggling to make ends meet, who finds somebody else's handbag unattended whilst clearing tables at a restaurant, with some folds of cash clearly visible at the top of the bag. It is here, again, proposed that in the face of temptation to take the money or otherwise pick up the bag unmolested and find its owner, the waitress's brain prepares the motor actions required for both responses, and weighs the relative costs of each in her final decision. Perhaps, as she glances around, she notices several people nearby and concludes that the risk is too high; or, instead, she may find the restaurant empty and, at the same time, remember the overdue mortgage payments that she cannot afford. Again, in such a state of preparation for either alternative, it becomes far easier to appreciate how and why the decisions of even ordinarily lawful individuals might tip into criminality when the correct constellation of circumstances and opportunity collide.

Such examples are not offered to provide absolution to the hypothetical individuals involved; provided that they possess the capacities for responsibility developed in Part Two of this thesis, those individuals would be responsible for any resultant crimes. Rather, the examples are to illustrate the implications of the aforementioned research and the brain's state of preparedness for multiple potential actions when faced with any given situation. If any such action under rapid consideration is criminal – as even the most conscientious and law-abiding individual might include amongst their options when the right circumstances and opportunities provide – and the mental representation of that criminal action is competing against other options in the brain for ultimate execution, it becomes considerably easier to understand how ordinary people can slip into criminality.

This is all the more pertinent when it is appreciated how the brain processes the *what* and *how* decision components automatically, *i.e.,* without the *necessary* intervention of

consciousness. Consequently, the facts that a brain considers a potentially criminal action amongst the available options for a given decision, prepares the motor actions necessary to implement that criminal action, and ultimately selects that criminal action, may each occur automatically and without realistic opportunity for an individual to consciously intervene. This seems inimical to the current approach to legal responsibility, which places central importance on *mens rea* on the premise that proof of certain subjective states of mind (*e.g.,* intention, recklessness, dishonesty, *etc.*) denotes that a given criminal decision has been reached freely, deliberately and consciously by the accused.

## *4.2. Volition and Agency*

Where the previous section discussed close connections between the action selection (*what*) and action preparation (*how*) components of a decision, the present section proceeds to explore how these elements of action selection and planning may combine together to further produce online control over our actions in motion and the accompanying sense of volition or agency that people experience. This is explained through the comparator model for action control and agency, which describes the combination of both prospective processes involved in the selection of actions and the predictive planning of their necessary physical movements and outcomes, and retrospective processes which compare predicted outcomes of actions with sensory feedback.

In broad terms, the comparator model proposes that the initiation of an action begins with the underlying goal which that action is intended to pursue. The accompanying plan computed to achieve that goal produces both the motor commands required to initiate and drive action, and an "efference copy" of these motor commands which proceeds to a forward modelling system to produce a prediction of the expected sensory consequences of that motor action being performed. This prediction is then compared with the actual sensory feedback informing the brain of both the ongoing motor actions and their effects on the environment.

Haggard describes three ways in which this comparison of the brain's prediction and sensory feedback is then used: (1) it provides the mechanism of comparing sensory feedback with expectations necessary to adjust current motor commands when there is a prediction error, thus providing online (and automatic) self-control over actions; (2) where there is no prediction error (*i.e.,* the prediction from the forward modelling system matches sensory feedback indicating that the motor action was completed successfully), to attribute agency over the individual's actions and their impact on the environment; and (3) to provide the conscious perception of self-generated actions and their predictable effects (see ***figure h***, below).[21]

The comparator model combines both prospective and retrospective processes involved in producing the sense of volition or agency. Again, in broad terms, prospective processes are those involved with the selection of an action and the prediction of its outcomes, whilst retrospective processes are engaged in comparing outcomes with their associated predictions and ascribing agency according to a lack of prediction error. This model is explored in further detail presently.

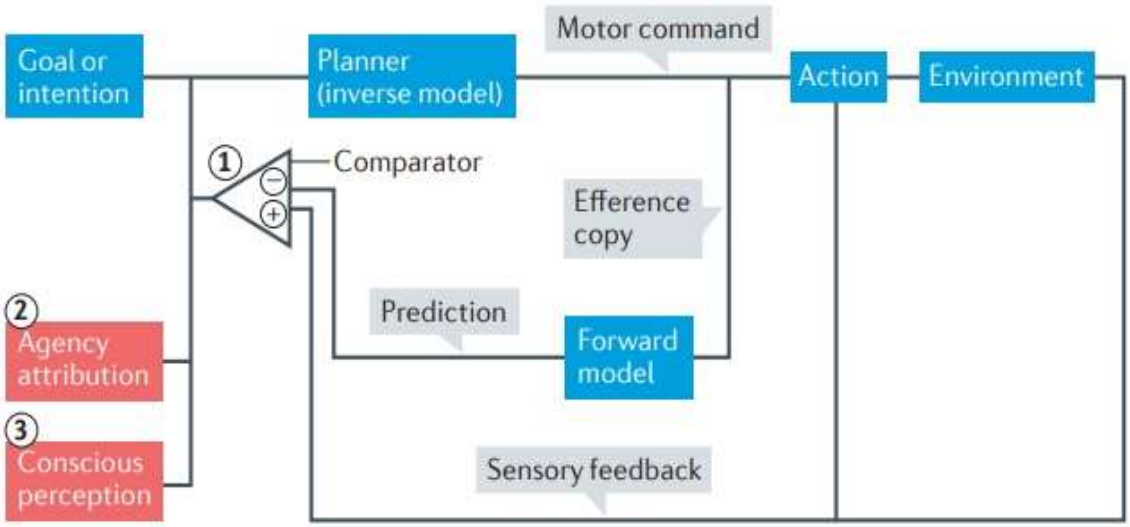


***Fig. h – The comparator model for control of action and sense of agency.***[22]

[21] Patrick Haggard, 'Sense of agency in the human brain' (2017) 18(4) *Nature Review Neuroscience* 196, 202.
[22] *Ibid.*

### *4.2.1. Prospective Processes*

One apparently key aspect of the sense of agency is 'some internal state of volition, conation or "urge"' which distinguishes voluntary from involuntary movements, the latter being such as produced by spasm or reflex *etc.*[23] Following from the seminal work of Kornhuber, Deecke *et. al.*[24] and Libet *et. al.*,[25] discussed in chapter five of this thesis, below, the Bereitschaftspotential or "Readiness Potential" ('RP') is classically associated with volitional movement. The RP refers to the slow negative electroencephalographic ('EEG') activity that can be reliably recorded starting in the SMA and pre-SMA approximately two seconds prior to the onset of volitional or intentional movements, but not before involuntary movements. Traditionally, the RP has been reasoned to reflected the generation of volition or an "urge" to act experienced before a person moves, whereas the RP is more recently reasoned to reflect motor preparation and conscious attention to a corresponding intention to move.

Haggard proposes that the 'cognitive preparation that precedes voluntary action may also contribute to sense of agency over an outcome.'[26] In support, he cites recent research by Jo, Wittmann, Hinterberger and Schmidt[27] which combines investigation of the RP alongside intentional binding, a phenomenon that is used as an implicit measure of the sense of agency. Intentional binding, discussed in more detail below, refers to a perceived shift in time between volitional actions and their intended effects – *i.e.,* people perceive the timing of intentional actions and their effects as being shifted closer together, whilst people perceive the timing of involuntary actions and effects as being further apart.[28]

---

[23] *Ibid*., 199.

[24] Hans H. Kornhuber and Lüder Deecke, 'Hirnpotentialänderungen bei willkürbewegungen und passiven bewegungen des menschen: Bereitschaftspotential und reafferente potentiale' (1965) 284(1) *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere* 1.

[25] Benjamin Libet, Elwood W. Wright Jr and Curtis A. Gleason, 'Preparation- or intention-to-act, in relation to pre-event potentials recorded at the vertex' (1983a) 56(4) *Electroencephalography and Clinical Neurophysiology* 367; Benjamin Libet, Elwood W. Wright Jr, Curtis A. Gleason and Dennis K. Pearl, 'Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act' (1983b) 106(3) *Brain: A Journal of Neurology* 623

[26] Haggard (2017), 200.

[27] Han-Gue Jo, Marc Wittmann, Thilo Hinterberger and Stefan Schmidt, 'The readiness potential reflects intentional binding' (2014) 8 *Frontiers in Human Neuroscience* 421.

[28] See Patrick Haggard, Sam Clark and Jeri Kalogeras, 'Voluntary action and conscious awareness' (2002) 5(4) *Nature Neuroscience* 382; Patrick Haggard and Sam Clark, 'Intentional action: Conscious experience

Jo *et. al.* took EEG recordings whilst subjects watched an analogue, single-handed clockface with a revolution period of 2,550 milliseconds, and reported their subjective timing of certain events over four conditions. In condition *baseline-M*, subjects performed a voluntary button press at a time of their choosing after at least one revolution of the clock, and then reported the position of the clock when they first started to move their finger. The *operant-M* condition was identical to *baseline-M* except for the addition of a tone presented 250 milliseconds after the button press. The *operant-T* condition was the same as *operant-M*, except subjects were instead asked to report the time that the tone played. And in the *baseline-T* condition the subjects did not perform any button press and reported the timing of a randomly presented tone.[29]

The intentional binding effect was displayed in the *operant-T* condition for 81% of subjects – that is, in the condition where the subjects' button press caused the subsequent tone, a majority of subjects experienced the timing of the tone as occurring earlier than it actually did.[30] When combined with the EEG data, a significant correlation was found between the early RP component and the *operant-T* condition such that the greater the early RP amplitude preceding button press, the larger the shift backwards in perceived time of the resultant tone.[31] The fact that this occurred only in the *operant-T* condition is interpreted as 'demonstrating that the early neural activity prior to movement plays a significant role in the consequent effect especially with respect to the sense of agency.'[32] As the RP is in turn associated with volition or attention thereto, these results point towards the involvement of this experience of volition or intention to act in the subsequent sense of agency over the effects of actions.

Relatedly, Haggard proposes that the processes of action selection – *i.e.,* choosing *what* to do – is another factor which prospectively contributes to the resultant sense of agency

and neural prediction' (2003) 12(4) *Consciousness and Cognition* 695; James W. Moore and Sukhvinder S. Obhi, 'Intentional binding and the sense of agency: A review' (2012) 21(1) *Consciousness and Cognition* 546.

[29] Jo *et. al.* (2014), 424.

[30] *Ibid*.

[31] *Ibid.,* 425.

[32] *Ibid.,* 427.

over actions.[33] A study by Barlas and Obhi[34] again explored the phenomenon of intentional binding, investigating in particular whether or not the effect varied when subjects could select from different numbers of alternative responses (either one, three or seven alternatives from which to choose). The results showed that the intentional binding effect was at its smallest when subjects had only one alternative to choose from or, effectively, no choice at all; here, only the perception of the timing of the subjects' choice of button press shifted in the correct direction towards the timing of the resultant tone. Subsequently, as the number of possible alternatives increased to three and then seven options, the intentional binding effect increased parametrically, with both the perception of the timing of the button press and the resultant tone shifting towards one another. This suggests that a 'high degree of choice is associated with greater action-effect binding than lower degrees of choice.'[35]

Khalighinejad, Di Costa and Haggard[36] present a meta-analysis over seven experiments involving 100 subjects. The experiments utilised transcranial direct current stimulation ('tDCS'), which is a non-invasive form a brain stimulation achieved through delivering a weak electrical current through two electrodes placed over the head. Across the seven experiments, tDCS was targeted towards the dlPFC, an area that has previously been strongly associated with the selection between, and monitoring of, endogenous voluntary actions. The various experiments then investigated whether the targeted tDCS would impact upon the intentional binding effect whilst subjects chose between multiple actions in various different conditions, including both endogenously selected and instructed actions with both the same and different outcome identities.[37] The parameters of each experiment thus varied according to action selection and action outcome.

Primary analysis of each experiment showed a mixed effect of anodal tDCS to the dlPFC, sometimes increasing the intentional binding effect (and action binding specifically) but

[33] Haggard (2017), 200.
[34] Zeynep Barlas and Sukhvinder S. Obhi, 'Freedom, choice, and the sense of agency' (2013) 7 *Frontiers in Human Neuroscience* 514.
[35] *Ibid*., 518.
[36] Nima Khalighinejad, Steven Di Costa and Patrick Haggard, 'Endogenous action selection processes in dorsolateral prefrontal cortex contribute to sense of agency: A meta-analysis of tDCS studies of "intentional binding"' (2016) 9(3) *Brain Stimulation* 372.
[37] *Ibid*., 373.

sometimes not. Meta-analysis of the results found moderate heterogeneity across the seven experiments, with 71% of the variability explained by classifying the studies according to whether the subjects' action was selected endogenously or instructed. Within the experiments where subjects selected their actions, tDCS had a modest but significant effect upon subsequent intentional binding to the subjects' selected action. More importantly for present purposes, the meta-analysis 'suggests a causal role for frontal action selection signals in prospective sense of agency'[38] – *i.e.,* the process of selecting *what* to do itself contributes to the resultant sense of agency over actions to implement that selection.

#### *4.2.2. Retrospective Processes*

Computational models of motor control have been used to describe and explain how the brain controls motor actions in the moment that they are being carried out, across a range of actions and sensory modalities. For example, when the eyes move to shift gaze from one position to another, a motor command must be created by the brain in order to direct the eye muscles to their new position. Concurrently, an "efference copy" of this motor command is created which produces a prediction of where that motor command should direct the eyes to gaze. The difference between the predicted movement and the sensory input from the actual movement can then be compared to ensure that the motor action has been completed accurately, and the eyes are correctly directed towards their new intended gaze.[39]

Similar processes are thought to be involved when controlling actions in other sensory modalities. For example, if you move your hand to scratch an itch on your arm, the brain expects to receive a sensory signal matching the timing, motion and frequency of the scratching movement, and compares the subsequent scratching sensation with those expectations to monitor and control the action producing it. Notably in this example, the initial itching sensation would not be accompanied by any efference copy and was

[38] *Ibid.,* 378; see also Valerian Chambon, Dorit Wenke, Stephen M. Fleming, Wolfgang Prinz and Patrick Haggard, 'An online neural substrate for sense of agency' (2013) 23(5) *Cerebral Cortex* 1031.

[39] See David S. Zee and Aasef G. Shaikh, 'The neurology of eye movements: From control systems to genetics to ion channels to targeted pharmacotherapy' in Werner J. S. and Chalupa L. M. (eds.), *The New Visual Neurosciences* (Massachusetts Institute of Technology 2014), 978.

'therefore unexpected by the brain.'[40] It is postulated that the central nervous system contains the forward modelling system which creates efference copies for, monitors, and controls motor actions.[41]

It has since been proposed that the aforementioned computational model for motor control may also provide the key retrospective mechanism which contributes to the sense of agency. Experiments conducted by Weiskrantz, Elliott and Darlington,[42] and separately by Claxton[43] explored the phenomenon that people are seemingly unable to tickle themselves or, at least, produce a considerably lesser sensation than when being tickled by another person. Weiskrantz *et. al*. utilised apparatus with which the subjects' feet were tickled with a weighted plastic pointer resting on a pivot and controlled by a handle. Subjects were then tickled across three different conditions: "active E" in which the experimenter controlled the handle to administer the tickling sensation; "active S" in which the subject controlled the handle to self-administer the tickling sensation; and "passive" in which the subject held the handle of the apparatus but allowed their own arm to be passively controlled and moved by the experimenter.

Thus, in the active E condition the subject neither produced motor commands nor received feedback stimulation from any motion of their arm, whilst both of these features would be present in the active S condition; and in the passive condition, the subjects would not produce any motor command (their arm being guided by the experimenter) but would produce sensory feedback from the motion of their own arm. Weiskrantz *et. al*. found that subjects' tickling sensation was strongest in the active E condition and weakest in the active S condition, with the passive condition producing weaker sensations than active E, but not so weak as active S. Considering the information available to the brain in each condition, they hypothesised that the cancellation of the tickling sensation could not be entirely based on the presence or absence of brain activity commanding the motion

[40] Elizabeth A. Styles, *Attention, Perception and Memory: An Integrated Introduction* (Psychology Press 2005), 165 – 166.
[41] Daniel M. Wolpert, 'Computational approaches to motor control' (1997) 1(6) *Trends in Cognitive Sciences* 209; Daniel M. Wolpert, R. Chris Miall and Mitsuo Kawato, 'Internal models in the cerebellum' (1998) 2(9) *Trends in Cognitive Sciences* 338.
[42] Lawrence Weiskrantz, John Elliott and Cyril D. Darlington, 'Preliminary observations on tickling oneself' (1971) 230(5296) *Nature* 598.
[43] Guy Claxton, 'Why can't we tickle ourselves' (1975) 41(1) *Perceptual and Motor Skills* 335.

of the subjects' arms – 'were that the case, both a self-administered tickle and the passive arm movement condition would be the same.'[44]

In Claxton's experiment, tickling sensations were delivered manually using a feather in one of four conditions: a) where the subjects' eyes were closed and the tickle administered by the experimenter, thus largely eliminating the possibility for the subject to predict the timing and location of the sensation; b) with the subjects' eyes open and the tickle administered by the administrator, such that the sensation still was not self-administered but became predictable; c) with the subjects' own arm passively administering the tickle whilst being controlled by the experimenter; and d) where the subject self-administered the tickle sensation. The strongest tickling sensation was produced in condition a) when the subjects' eyes were closed and they were tickled by the experimenter. Claxton proposes that subjects in this condition did not have the opportunity to "steel themselves"; more formally, he suggests that there may exist a high-level (and *perhaps* conscious) 'ability to control the perceived magnitude of sensation.'[45] The next strongest sensation was produced in condition b) when the subjects' eyes were open but they were still being tickled by the experimenter.

The tickling sensation was third strongest in condition c), during which subjects would receive feedback from the motion of their own arms, but which were being caused to move by the experimenter. The weakest tickling sensation was elicited in condition d) when subjects attempted to tickle themselves voluntarily, however the difference between conditions c) and d) was not statistically significant. By contrasting conditions a) and b) with c) and d) in particular, Claxton suggests that both the predictability of a tickling sensation and sensorimotor feedback from movements of the subjects' arms had a significant effect upon the resultant experience of a tickling sensation.

A classic paper by Blakemore, Wolpert and Frith[46] modifies the paradigm by Weiskrantz *et. al*. to examine the perceptual effects of variations between self-generated movements

[44] Weiskrantz, Elliott and Darlington (1971), 598.
[45] Claxton (1975), 337.
[46] Sarah-Jayne Blakemore, Daniel Wolpert and Chris Frith, 'Why can't you tickle yourself?' (2000) 11(11) *Neuroreport* R11.

and their sensory results, by parametrically altering changes in delay and trajectory between subjects' own movements and their resultant perceptual stimulation. Robotic apparatus was created which produced tactile stimulation of the subjects' right hand by applying a sinusoidal motion with a piece of soft foam. This sensation was created under a range of different conditions; in one condition, the sensation was produced by the robotics with no relation to any movements from the subjects. In subsequent conditions, subjects were instructed to make the sinusoidal motions with their left hand, which were then mirrored by the robot to produce the actual sensations on the subjects' right hand. In a self-produced condition, the stimulation delivered by the robotic apparatus corresponded exactly with the subjects' own motions. In subsequent conditions, delays of 100, 200 and 300 milliseconds were interposed between the motion of the subjects' left hand and the robotic stimulation to their right hand; and in further conditions, trajectory rotations of 30°, 60° and 90° were interposed between the direction of the subjects' left-hand motions and the robotic stimulation of their right hand.

Thus, by introducing varied degrees of delay or trajectory rotation between the subjects' own motions and the sensations they produced, the brain's sensory predictions of the results of those motions became less accurate, with three possible effects being produced. First, if attenuation to the tactile stimulation was caused by 'general movement-induced sensory gating', then it would be expected that sensory attenuation would occur equally under the various delay and trajectory rotation conditions as the subjects' sensation of moving their left arm remains substantially the same. Second, if attenuation to the sensory stimulus were dependent upon an entirely accurate prediction of that sensation, no attenuation would be expected under the delay and trajectory rotation conditions. Third, sensory attenuation may occur in proportion to the accuracy of prediction, in which case the intensity of the sensation would increase as the delay or trajectory rotation increased also.[47]

Subjects reported the self-produced sensation to be significantly less intense and tickly than the identical sensation produced by the robotic arm alone. Moreover, as the delay between the subjects' arm motion and the resultant stimulation increased from 0 to 200

[47] *Ibid.*, R13.

milliseconds, and as the trajectory rotation was increased between 0° and 90°, subjects reported a corresponding increase in the intensity of the tickling sensation. Blakemore, Wolpert and Frith submit that these results 'support the hypothesis that the perceptual attenuation of self-produced tactile stimulation is due to precise sensory predictions, rather than a movement-induced non-specific attenuation of all sensory signals.'[48]

As Haggard explains, these experiments provide behavioural evidence for the key retrospective process involved in producing a sense of agency. Specifically, computational models of motor control assert that motor commands are produced alongside an efference copy of those commands which pass to a forward (or internal predictive) model. That forward model is subsequently compared with resultant sensory feedback from the motor action to produce a prediction error, from which the brain can deduce whether or not the motor command has been carried out properly according to predictions and whether the motor action needs to be modified if an error has occurred, which would produce a mismatch between the forward model and the actual sensory feedback received.[49]

The same process of comparing sensory feedback with an efference copy of motor commands is hypothesised to be similarly involved in producing the experience of agency over actions. Haggard writes, 'if an event is caused by one's own action (and if the internal predictive model is correct), the actual feedback corresponds exactly to the prediction, and the result of the comparison is zero; otherwise, the result is a non-zero prediction error.'[50] Thus, people experience the sense of agency 'over events that can be predicted given their motor commands.'[51] In further support, Haggard cites an experiment by Farrer *et. al.*[52] which involved introducing dissonance between subjects' actual motor movements and sensory feedback they viewed on a computer screen, similarly to the work by Blakemore, Wolpert and Frith. Subjects were asked to attribute motions on a computer screen either to their own movements or those of the experimenter when, in fact, all of

[48] *Ibid*.
[49] Haggard (2017), 201.
[50] *Ibid*.
[51] *Ibid*.
[52] Chlöe Farrer, Mathilde Bouchereau, Marc Jeannerod and Nicolas Franck, 'Effect of distorted visual feedback on the sense of agency' (2008) 19(1-2) *Behavioural Neurology* 53.

the computer motions were caused by their own movements. Nonetheless, subjects attributed movements to another when there was a high *spatial* discordance between their own movements and the sensory feedback received, highlighting direction of movements as a 'cardinal feature in action attribution.'[53] Further supporting experiments have similarly shown the importance of temporal cues with the 'time of initiating an action allow[ing] a precise prediction about the time of the outcome.'[54]

#### *4.2.3. Summary Discussion on Volition and Agency*

Three broader points may be extrapolated and discussed from the above exploration of features of action planning, agency, and the *how* component of decision-making. First, the present chapter began by highlighting the apparent functional proximity between the *what* and *how* components – *i.e.,* as part of the process of deciding *what* to do, the brain concurrently prepares actions plans for the various different options under consideration. In a similar vein, both the *what* and *how* components appear to be similarly involved in producing the sense of agency over actions by providing prospective and retrospective processes respectively. Thus, the process of deciding *what* to do prospectively contributes to the sense of agency, whilst the production of an efference copy of motor plans and the comparison of predicted and actual outcomes contributes to the sense of agency retrospectively. Indeed, it is highly unlikely that any of the components of decision-making operate in isolation in the brain, as each must contribute to the execution of a final and complete decision. Nevertheless, the evidence considered in the present chapter demonstrates connections between the *what* and *how* components in particular.

Second and relatedly, the present chapter has highlighted the connection between the *how* component of decision-making and the sense of agency. The preceding section explored in particular the expansion of computational models of motor control to the production of a sense of agency, in particular through the retrospective process of comparing forward models of motor plans with actual sensory feedback. This connection is further reinforced

[53] *Ibid*., 53.

[54] Haggard (2017), 202; citing Chess Stetson, Xu Cui, P. Read Montague and David M. Eagleman, 'Motor-sensory recalibration leads to an illusory reversal of action and sensation' (2006) 51(5) *Neuron* 651; Eamonn Walsh and Patrick Haggard, 'Action, prediction, and temporal awareness' (2013) 142(2) *Acta Psychologica* 220.

when considering the areas of the brain that have been indicated in both motor planning and the attribution of agency. Section 4.1 of this thesis, above, identified a number of such areas involved in the planning of endogenous motor actions, including (but not limited to) the SPL, IPS and SMG.

Farrer *et. al.*[55] present a modified version of the paradigm by Blakemore, Wolpert and Frith in which discordance between subjects' motor actions and their resultant effects on a computer screen were introduced by rotating the trajectory of those effects from the subjects' actions, whilst subjects underwent scans using positron emission tomography ('PET'). The key results found that brain activity increased in the inferior parietal lobule ('IPL') and, more specifically, the SMG and angular gyrus ('AG') 'as a function of the degree of discordance between the executed and seen movements', with this increase in activity corresponding to a decrease in the subjects' sense of agency over those movements. Concurrently, activity in the posterior insula decreased corresponding with an increase in subjects' sense of agency. Thus, the authors conclude that activity in the inferior parietal cortex 'relates to the feeling of loss of agency associated with the discrepancy between intended actions and sensory feedback', whilst activity in the posterior insula relates to the feeling of agency over actions.

What is particularly illuminating is the topographical relationship between those parietal areas engaged in action planning and the sense of agency. The IPL (associated with a loss of the sense of agency) sits directly beneath the SPL (associated with action planning), separated by the IPS (also associated with action planning). Furthermore, the SMG (associated with *both* action planning and loss of the sense of agency) sits adjacent to the AG (also associated with loss of the sense of agency). Thus, the discovery of regions of the brain engaged in motor planning and sense of agency being situated adjacently (SMG to AG; SPL to IPL) or overlapping (SMG) provides further compelling evidence that processes engaged in the planning of motor actions also contribute to the phenomenon of agency over those actions. This topographical connection between brain networks engaged in both motor planning and sense of agency is further bolstered by studies

[55] Chlöe Farrer, Nicolas Franck, Nicolas Georgieff, Chris Frith, Jean Decety and Marc Jeannerod, 'Modulating the experience of agency: A positron emission tomography study' (2003) 18(2) *Neuroimage* 324; see also Chlöe Farrer and Chris Frith, 'Experiencing oneself vs another person as being the cause of an action: The neural correlates of the experience of agency' (2002) 15(3) *Neuroimage* 596.

revealing distortions in the attribution of agency in patients who have suffered lesions in these regions of the brain or where activity is artificially depressed by way of TMS.[56]

Third and finally, as with other components of decision-making, there is mounting evidence that pathologies within the networks associated with the *how* component of decision-making can manifest as mental illnesses with significant effects on people's behaviour. In brief, impairments, deficits and distortions in sense of agency have been associated with schizophrenia,[57] obsessive-compulsive disorder,[58] autism spectrum disorder,[59] psychogenic movement disorder,[60] and other neurological disorders such as anarchic and alien hand disorders.[61] However, it is schizophrenia patients who arguably provide the '"pathophysiology model" for agency processing, *i.e.,* they provide a window to the processes underlying one's self-attribution of actions.'[62]

Schizophrenia is typically characterised by symptoms which 'testify to an impairment in self-attributing their own thoughts or actions.'[63] Thus, so-called "first rank symptoms" such as verbal hallucinations, thought insertion or removal, and delusions of control by

[56] For example, see Angela Sirigu, Elena Daprati, Pascale Pradat-Diehl, Nicolas Franck and Marc Jeannerod, 'Perception of self-generated movement following left parietal lesion' (1999) 122(10) *Brain* 1867; see also Penny A. MacDonald and Tomás Paus, 'The role of parietal cortex in awareness of self-generated movements: A transcranial magnetic stimulation study' (2003) 13(9) *Cerebral Cortex* 962; Mariella Pazzaglia and Giulia Galli, 'Loss of agency in apraxia' (2014) 8 *Frontiers in Human Neuroscience* 751; Nima Khalighinejad and Patrick Haggard, 'Modulating human sense of agency with non-invasive brain stimulation' (2015) 69 *Cortex* 93.

[57] Matthis Synofzik, Gottfried Vosgerau and Martin Voss, 'The experience of agency: An interplay between prediction and postdiction' (2013) 4(127) *Frontiers in Psychology* 1.

[58] Antje Gentsch, Simone Schütz-Bosbach, Tanja Endrass and Norbert Kathmann, 'Dysfunctional forward model mechanisms and aberrant sense of agency in obsessive-compulsive disorder' (2012) 71(7) *Biological Psychiatry* 652.

[59] Emma Gowen and Antonia Hamilton, 'Motor abilities in autism: A review using a computational context' (2013) 43(2) *Journal of Autism and Developmental Disorders* 323.

[60] Isabel Pareés, Harriet Brown, Atsuo Nuruki, Rick A. Adams, Marco Davare, Kailash P. Bhatia, Karl Friston and Mark J. Edwards, 'Loss of sensory attenuation in patients with functional (psychogenic) movement disorders' (2014) 137(11) *Brain* 2916.

[61] James W. Moore and Paul C. Fletcher, 'Sense of agency in health and disease: A review of cue integration approaches' (2012) 21(1) *Consciousness and Cognition* 59.

[62] Synofzik, Vosgerau and Voss (2013), 5; see also Lukas Uhlmann, Mareike Pazen, Bianca M. van Kemenade, Tilo Kircher and Benjamin Straube, 'Neural correlates of self-other distinction in patients with schizophrenia spectrum disorders: The role of agency and hand identity' (2021) 47(5) *Schizophrenia Bulletin* 1399; Sukhwinder S. Shergill, Gabrielle Samson, Paul M. Bays, Chris D. Frith and Daniel M. Wolpert, 'Evidence for sensory prediction deficits in schizophrenia' (2005) 162(12) *American Journal of Psychiatry* 2384.

[63] Marc Jeannerod, 'The sense of agency and its disturbances in schizophrenia: A reappraisal' (2008) 192(3) *Experimental Brain Research* 527, 530.

alien entities each involve experiences of not being in control of oneself and instead being controlled by external agents. Such misattribution can also operate in the opposite direction, *i.e.,* when patients over-attribute effects in the world to their own actions, such as believing that they can control the thoughts and behaviour of others. The 'current explanation for the first rank symptoms, as proposed by Feinberg[64] and Frith,[65] is that schizophrenic patients lose the normal ability to monitor one's self-willed intentions and actions.'[66]

Haggard *et. al.*[67] adapted the experimental paradigm used by Jo *et. al.*[68] and described in section 4.2.1 of this thesis, above, administering the key-press exercise to eight schizophrenic patients alongside matched healthy controls and comparing the effects of intentional binding between the two groups. The results found that schizophrenic patients exhibited a significantly stronger intentional binding effect than healthy controls, suggesting that patients might over-associate their own actions with subsequent events. However, this result was somewhat curious considering that patients more commonly cite feelings of being out of control and often under the influence of some outside force or agent.

Voss *et. al*.[69] repeated a similar experiment with 24 schizophrenic patients and matched controls which was adapted to 'isolate the respective predictive and retrospective contributions to actions experience.'[70] In particular, the probability with which the subjects' key press would elicit a resultant sound was varied such that subjects could predict that their actions would produce the sound on some trials – (engaging predictive processes in intentional binding) – whereas the sound would be unpredictable in other

[64] Irwin Feinberg, 'Efference copy and corollary discharge: Implications for thinking and its disorders' (1978) 4(4) *Schizophrenia Bulletin* 636.
[65] Christopher D. Frith, *The Cognitive Neuropsychology of Schizophrenia* (Lawrence Erlbaum Associates 1992).
[66] Jeannerod (2008), 530.
[67] Patrick Haggard, Flavie Martin, Marisa Taylor-Clarke, Marc Jeanerod and Nicolas Franck, 'Awareness of action in schizophrenia' (2003) 14(7) *NeuroReport* 1081.
[68] Jo, Wittmann, Hinterberger and Schmidt (2014).
[69] Martin Voss, James Moore, Marta Hauser, Juergen Gallinat, Andreas Heinz and Patrick Haggard, 'Altered awareness of action in schizophrenia: A specific deficit in predicting action consequences' (2010) 133(10) *Brain* 3104.
[70] *Ibid.,* 3106.

trials – (engaging retrospective processes in intentional binding).[71] Concurring with the results of Haggard *et. al.*, above, schizophrenia patients exhibited significantly greater binding effects than healthy controls; however, most such binding occurred in patients when a tone occurred, suggesting a 'greater influence of sensory driven, retrospective processes on action awareness in patients.'[72] For controls, the greatest binding effect occurred in the condition when there was a higher probability of causing the tone by pressing the button, suggesting a 'greater influence of predictive processes on action awareness.'[73]

The results suggest that patients with schizophrenia experience a deficit in predicting the consequences of their own action plans, and a corresponding exaggerated reliance on making retrospective connections between phenomena in the world and their own actions. Furthermore, the results showed that the 'schizophrenic deficit in predicting the relation between action and effect was strongly correlated with severity of positive psychotic symptoms, specifically delusions and hallucinations.'[74] This latter point is supported by research from Krugwasser *et. al.*[75] which suggests that patients suffering from other forms of psychosis also exhibit a 'severely reduced ability for discriminating their actions… [and] do not show proper metacognitive insight into this deficit.'[76] Sense of agency, therefore, appears crucially important for distinguishing the self from the environment, and *correctly* attributing phenomena in the environment to one's own actions, the actions of others, or happenstance.[77]

[71] See further James W. Moore and Patrick Haggard, 'Awareness of action: Inference and prediction' (2008) 17(1) *Consciousness and Cognition* 136.
[72] Voss *et. al.* (2010), 3108.
[73] *Ibid.*
[74] *Ibid.,* 3104.
[75] Amit Regev Krugwasser, Yonatan Stern, Nathanb Faivre, Eiran Vadim Harel and Roy Salomon, 'Impair sense of agency and associated confidence in psychosis' (2022) 8(32) *Schizophrenia* 1.
[76] *Ibid.,* 1; see also Marta Hauser, Guenther Knoblich, Bruno H. Repp, Marion Lautenschlager, Juergen Gallinat, Andreas Heinz and Martin Voss, 'Altered sense of agency in schizophrenia and the putative psychotic prodrome' (2011) 186(2-3) *Psychiatry Research* 170.
[77] See further Jean-Rémy Martin, 'Experiences of activity and causality in schizophrenia: When predictive deficits lead to a retrospective over-binding' (2013) 22(4) *Consciousness and Cognition* 1361; Synofzik, Vosgerau and Voss (2013), 5 – 6.

## *4.3. From Action Planning and Agency to Legal Responsibility*

The broader importance of this research to the present thesis follows: as mentioned above, schizophrenia offers but one example out of many psychiatric disorders which can have a significant impact on behaviour and are linked to impairments, deficits and distortions in sense of agency, which is itself a product of the brain mechanisms governing the *how* component of decision-making. Meanwhile, the disproportionate prevalence of convicts with such associated psychiatric disorders within the prison population is not only well documented in the United Kingdom but in jurisdictions around the world; untreated mental illnesses such as are related to deficiencies in the *how* processes of decision-making can readily deteriorate into decisions and behaviour that are criminal.

Against a reported prevalence of psychotic disorders amongst 0.7% of the adult population of England,[78] between 10% and 12% of the nation's prison population meets the criteria for having suffered from psychosis;[79] this is contrasted with a rate of between 3.6% and 3.9% for global prison populations.[80] More generally, it is estimated that between 17% and 25% of the UK population experience some form of mental health problems, whilst the Ministry of Justice estimates that more than half of prisoners have common psychiatric disorders (including depression, anxiety and post-traumatic stress disorder), with approximately 15% of prisoners having specialist mental health needs and around 2% suffering from serious or acute mental health problems.[81]

The types of mental health disorders associated with faults in the *how* component of decision-making and the sense of agency – in particular schizophrenia and other disorders manifesting in psychosis – have critical implications for holding those who suffer from such disorders as being responsible for their actions. Discussed in section 3.3, above, the existence of a defendant's subjective mental states – such as intention, recklessness and

[78] Public Health England, *Psychosis Data Report: Describing variation in numbers of people with psychosis and their access to care in England* (Crown Copyright 2016), 15.
[79] Paul Bebbington, Sharon Jakobowitz, Nigel McKenzie, Helen Killaspy, Rachel Iveson, Gary Duffield and Mark Kerr, 'Assessing needs for psychiatric treatment in prisoners: 1. Prevalence of disorder' (2017) 52(2) *Social Psychiatry and Psychiatric Epidemiology* 221.
[80] Seena Fazel and Katharina Seewald, 'Severe mental illness in 33,588 prisoners worldwide: Systematic review and meta-regression analysis' (2012) 200(5) *British Journal of Psychiatry* 364.
[81] House of Commons Committee of Public Accounts, *Mental Health in Prisons* (HC 400, Eight Report of Session 2017-19), 9.

dishonesty – is not itself necessarily probative that that criminal *mens rea* has been consciously and deliberately chosen by an individual, as is otherwise presumed by the law. The discussion in the present chapter of this thesis similarly concludes that the existence of such subjective mental states does not necessarily prove that an individual has understood the nature and consequences of their actions.

The first rank symptoms of schizophrenia – verbal hallucinations, thought insertion or removal, and delusions of control by alien entities – as well as megalomania – when patients believe that they can control the thoughts and behaviour of others – and more general symptoms of psychosis, are each understood as resulting from the misattribution of phenomena in the world to the patient's own actions or the actions of others. Put differently, these various symptoms appear to result (at least in part) from deficiencies in the *how* component of decision-making which contributes to the sense of agency, this appearing to be fundamental for distinguishing the self (and one's own deliberate actions) from others (and their actions) and the environment. Thus, deficiencies in the *how* component of decision-making, and sense of agency specifically, can directly impair an individual's capacity to appreciate the nature of their actions and the consequences they have in the world, as is typical in sufferers of schizophrenia and psychosis.

Consider the hypothetical where a patient suffering from psychosis incorrectly perceives another individual as being a violent assailant or, even more fantastically, some alien creature that is attacking. Responding to this perceived danger, the patient acts in "self-defence" and *intentionally* shoots his assailant with a gun, killing an otherwise innocent individual. There is little difficulty in concluding that the patient suffering from psychosis is not responsible for the actions of killing another person, which he genuinely perceived and believed to be an attacking alien. Yet, the patient would *prima facie* possess the requisite *mens rea* of intention to kill or cause grievous bodily harm – after all, the patient *intentionally* shot the gun to stop or kill his assailant.

Crucially, however, the hypothetical patient lacked the capacity to appreciate and understand the nature of his actions – *i.e.,* that he was actually shooting towards another human being – and the consequences of those actions – *i.e.,* that he would grievously

injure and quite possibly kill that other person and concurrently break the law. Impairments, deficits and distortions in the sense of agency, (which is itself a product of the *how* component of decision-making), can result in people committing criminal actions whilst in possession of the requisite subjective state of mind (*mens rea*), all the while lacking the capacity to understand the nature of those actions or their criminal consequences.

Chapter nine of the present thesis proposes that one of three necessary capacities for holding individuals responsible for their actions includes the capacity to appreciate the nature and consequences of their actions. When this capacity is impaired or absent altogether, individuals may readily become unable to understand that a particular decision to act will result in criminal consequences, or they may act upon their perception of phenomena erroneously misattributed either to their own actions or the actions of others. In such circumstances, it may be argued that people lack the broader ability to ensure that their actions comply with criminal proscriptions, as they cannot necessarily appreciate how or why a given action is itself criminal. Consequently, it becomes increasingly unreasonable to expect that such affected individuals should or even could conform their behaviour entirely with the criminal law.

Whilst the intervention of the law may be nonetheless necessary for the protection of individuals afflicted with certain mental illnesses and the wider society, both labelling and treating those individuals as "criminals" scarcely seems to be a just and fair response. Nor is the overrepresentation of such individuals within the prison system likely to be the most effective, let alone fair, means of securing the treatment and rehabilitation that they require. Further still, the analogy is readily drawn between the case of the patient whose criminal conduct was the demonstrable result of a brain tumour – presented in the introduction to this thesis – and cases of criminal behaviour resulting from defects in the decision-making mechanisms of the brain, presenting as various mental illnesses.

In both types of case, the "flaw" in an individual's decision-making processes clearly results from factors over which they have practically no influence or control – their very decision-making faculties are set up in such a way that all but guarantees their eventual

criminal conduct. Whilst, again, it may remain nonetheless necessary for the law to intervene for the purposes of securing the greater safety and security of both afflicted individuals and the wider society, it is incumbent upon the justice system to be more compassionate and rehabilitative in such cases. Chapter twelve of the present thesis explores this argument further and proposes relevant reforms to both the verdicts that may be rendered in these cases, and the theories of "punishment" that may fairly and legitimately be applied.

# 5. The *When* Component and Timing of Consciousness

> 'There is no absolute or free will in the mind; but the mind is determined to will this or that by a cause, which is also determined by another cause, and this in turn by another, and so on *ad inifinitum*.'
>
> - Baruch Spinoza, 1677.[1]

As Brass and Haggard discuss in their original proposal, the *when* component has arguably received the greatest attention in academic literature regarding intentional action, 'perhaps because it leads naturally to questions about free will and the causation of intentional actions.'[2] The authors continue, 'the onset time of brain processes culminating in intentional action has proved important in understand the flexibility and genesis of action, as well as providing a neuroscientific perspective on "free will".'[3] Such perspectives have been largely informed by research investigating the point in time during a decision-making process at which individuals become consciously aware of making a decision. This invites further questions concerning the very role that consciousness itself plays in decision-making – a recurrent theme throughout this chapter for which many crucial questions remain unanswered in the current state of the art.

As Blackmore and Troscianko explain, to question the role of consciousness in decision-making is not to cast doubt upon whether or not human beings are agents who make decisions, nor whether processes of emotion, thought and deliberation can be consciously engaged in making those decisions. Like all creatures, humans are biological beings which interact with the wider world; 'they respond to events, make intricate plans with many available options, and act accordingly, at least when not restrained or coerced.'[4]

---

[1] Baruch Spinoza, *Ethics: Demonstrated in Geometric Order* (Kisner M. J. (ed.), Cambridge University Press 2018), 85.
[2] Marcel Brass and Patrick Haggard, 'The what, when, whether model of intentional action' (2008) 14(4) *Neuroscientist* 319, 320.
[3] *Ibid*., 321.
[4] Susan Blackmore and Emily T. Troscianko, *Consciousness: An Introduction* (3rd ed. Routledge 2018), 222.

Moreover, as with other intelligent animals (albeit arguably to a greater extent), humans consider and compare different possible actions and their likely outcomes, whilst scientific pursuits such as the various branches of neuroscience and psychology allow us to 'look to see which parts of the brain and the rest of the body are involved in such decision-making and, in principle at least, trace how they lead to particular decisions and actions.'[5]

In exploring the role of consciousness, however, Blackmore and Troscianko question whether this is 'any different from exploring Google's search algorithms to see how it chose which list of links to show me when I asked it "what is consciousness?".'[6] The assumption here is that such search results are ultimately fully determined by the underlying algorithms which produce them, no matter how complicated they may be, whereas the contrary assumption is often made regarding decision-making in people. The question, therefore, is whether consciousness plays any more necessary a role in human decisions-making, and what are the implications of this question for concepts such as volitional control over decision-making and free will.

The aim of this chapter is to demonstrate how the conscious awareness of any given decision is preceded by unconscious cognitive processes which first result in that decision. It is submitted that the high degree of predictability that these cognitive processes indicate towards the outcome of decisions strongly suggests that a decision is first reached by the unconscious brain, after which individuals become consciously aware of what that decision is. If this ordering is correct, it implies a significantly reduced, if at all existent, ability for individuals to *consciously* control the decisions which they make, whereas the existence of direct conscious control over decisions and actions is otherwise presumed to exist as one of the components of the legal concept of volition.

[5] *Ibid*.
[6] *Ibid*.

## 5.1. ***The Initiation of Volitional Action***

From a series of seminal work beginning the 1960s, Kornhuber, Deecke, *et. al.*[7] are accredited with the discovery of the Bereitschaftspotential or "Readiness Potential" ('RP'), consisting of the slow negative electroencephalographic ('EEG') activity that can be reliably recorded, starting in the supplementary motor area ('SMA') and pre-SMA, approximately 2 seconds prior to the onset of volitional or intentional movements, but not before involuntary movements. A simple experimental paradigm consists of instructing subjects to voluntarily move a certain part of their body, on their own accord and at irregular intervals (*i.e.,* not simply following a regular pattern of behaviour), whilst EEG recordings were taken from the scalp. Recordings can be contrasted against subjects being caused to make involuntary movements, such as by pulling a string on the finger which causes it to flex, or inducing movement by transcranial magnetic stimulation ('TMS'). Experiments consistently show that for *voluntary movements only*, the RP begins approximately 2 seconds prior to movement with a slow negative slope in the pre-SMA (unlocalised) and in the SMA according to somatotopic organisation. At around 400 – 500 milliseconds prior to movement onset, a steeper negative slope is recorded in the primary motor cortex contralateral to movement.

Drawing from their extensive body of research, Kornhuber, Deecke, *et. al.* conclude that:

> '[T]he supplementary motor area [] is the central key structure transducing the will-to-move into effective actions. In other words, the SMA has a common starting function for various kinds of volitional actions…'[8]

[7] Hans H. Kornhuber and Lüder Deecke, 'Hirnpotentialänderungen bei willkürbewegungen und passiven bewegungen des menschen: Bereitschaftspotential und reafferente potentiale' (1965) 284(1) *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere* 1; Lüder Deecke, Peter Scheid and Hans H. Kornhuber, 'Distribution of readiness potential, pre-motion positivity, and motor potential of the human cerebral cortex preceding voluntary finger movements' (1969) 7(2) *Experimental Brain Research* 158; Lüder Deecke, Berta Grözinger and Hans H. Kornhuber, 'Voluntary finger movement in man: Cerebral potentials and theory' (1976) 23(2) *Biological Cybernetics* 99; Lüder Deecke and Hans H. Kornhuber, 'An electrical sign of participation of the mesial "supplementary" motor cortex in human voluntary finger movement' (1978) 159(2) *Brain Research* 473; Hans H. Kornhuber and Lüder Deecke, 'Readiness for movement – The bereitschaftspotential story' (1990) 33(4) *Current Contents Life Sciences* 14.

[8] Hans H. Kornhuber, Lüder Deecke, Wilfried Lang, Michael Lang and Anselm Kornhuber, 'Will, volitional action, attention and cerebral potentials in man: *Bereitschaftspotential* performance-related potentials,

Aside from the evidence of the RP which is only adduced in relation to volitional movements, there are a number of further functional and anatomical features of the SMA which support this conclusion. Kornhuber, Deecke *et. al.* continue, although the control of various different motor movements is widely decentralised in the human brain, the primary initiation of the RP in the SMA followed by a cascade to further motor cortices points towards the SMA acting as a common structure for the initiation of all voluntary movement – *i.e.,* the decision of *when* to act. As the authors note, this process requires input from the brain's motivational system; in this regard, the SMA receives input directly or indirectly (via the thalamus) from the hypothalamus, amygdala, inferior temporal cortex and prefrontal cortex, each of which are associated with producing various motivational states.[9] They add that patients with lesions in the SMA characteristically continue to experience the will to action and are able to select between alternative motives, 'but the transduction of their intentions into actions is disrupted.'[10] In a similar vein, selecting when to initiate movement requires that any necessary anticipatory mechanisms are sufficiently prepared, such as to adjust posture and balance in response to the anticipated outcome of motion. In this regard, again, the SMA receives input from associated brain regions in the sensorimotor cortex, from the cerebellum and basal ganglia and via the thalamus. This suggests that the SMA contains the requisite functional connections to other regions of the brain in order to receive the information necessary to decide when to initiate a planned movement.

More modern research continues to associate the RP with the experience of volition or the conscious intention to act, although this connection remains somewhat contentious, and it is unclear whether the RP could be causative of, or merely correlative with, intention. Shibasaki and Hellet write that the early, slow RP 'might reflect, physiologically, slowly increasing cortical excitability and, behaviorally, subconscious readiness for the forthcoming movement,' continuing, '[w]hether the late RP reflects

direction attention potential, EEG spectrum changes' in W. A. Hershberger (ed.), *Volitional Action: Vol. 62* (North Holland 1989), 118; see also Vinh T. Nguyen, Michael Breakspear and Ross Cunnington, 'Reciprocal interactions of the SMA and cingulate cortex sustain premovement activity for voluntary actions' (2014) 34(49) *Journal of Neuroscience* 16397.

[9] Kornhuber, Deecke, Lang, Lang and Kornhuber (1989), 119.

[10] *Ibid*.

*conscious* preparation for intended movement or not remains to be clarified.'[11] In this latter regard, Haggard and Eimer[12] produced a replication of Libet's work (discussed in the following section, below), and indeed found that the lateralised RP (and not the ipsilateral RP) was correlated with how early or late subjects reported being aware of an intention to move, supporting the aforementioned hypothesis. However, Haggard and Eimer's results later failed to replicate in work by Schelgel *et. al.*,[13] the latter arguing that the RP and lateralised RP both reflected processes independent of conscious will.

One particularly illuminating insight into the RP has been provided by studies of patients who experience physical tics as a result of Tourette's Syndrome; between 80%[14] and 90%[15] of patients experience a premonitory urge to perform a tic before doing so. Duggal and Nizamie[16] studied three patients who did experience the premonitory urge, and the RP was recorded in all patients preceding their tics. So much might be expected if indeed the RP is related to volitional urge. What renders the finding more intriguing is comparison with an earlier study by Karp *et. al.*[17] involving five patients with Tourette's syndrome, of whom only one reported having premonitory urges. This study recorded the RP preceding tics in two of the five subjects, one being the patient who experienced premonitory urges (albeit the RP was also recorded in one other patient who did not experience such urges preceding tics). The finding across both studies that the RP is more likely to be elicited preceding the tics of subjects who also experienced premonitory urges

[11] Hiroshi Shibasaki and Mark Hallet, 'What is the bereitschaftspotential?' (2006) 117(11) *Clinical Neurophysiology* 2341, 2341.
[12] Patrick Haggard and Martin Eimer, 'On the relation between brain potentials and the awareness of voluntary movements' (1999) 126(1) *Experimental Brain Research* 128.
[13] Alexander Schlegel, Prescott Alexander, Walter Sinnott-Armstrong, Adina Roskies, Peter U. Tse and Thalia Wheatley, 'Barking up the wrong free: Readiness potentials reflect processes independent of conscious will' (2013) 229(3) *Experimental Brain Research* 329.
[14] Amy J. Cohen and James F. Leckman, 'Sensory phenomena associated with Gilles de la Tourette's syndrome' (1992) 53(9) *Journal of Clinical Psychiatry* 319.
[15] David C. Houghton, Matthew R. Capriotti, Christine A. Conelea and Douglas W. Woods, 'Sensory phenomena in Tourette syndrome: Their role in symptom formation and treatment' (2014) 1(4) *Current Developmental Disorders Reports* 245.
[16] H. S. Duggal and S. Haque Nizamie, 'Bereitschaftspotential in tic disorders: A preliminary observation' (2002) 50(4) *Neurology India* 487.
[17] Barbara Illowsky Karp, Simone Porter, Camilo Toro and Mark Hallett, 'Simple motor tics may be preceded by a premotor potential' (1996) 61(1) *Journal of Neurology, Neurosurgery and Psychiatry* 103.

lends further evidence towards the involvement of this element in the experience of volition.

Some of the most recent research surrounding the bereitschaftspotential and volition has suggested that the RP may in fact be more reflective of *attention* to, or *awareness* of, volition as opposed to necessarily reflecting volition itself. Takashima *et. al.*[18] deploy a paradigm within which subjects perform a spontaneous button press whilst recordings are taken by EEG. The subjects are distracted in one condition, for example, by performing a mental imagery task before the button press, whilst subjects in the second condition were instructed to pay particular attention to their intention to act before and during the button press and were not otherwise distracted. Consistently across three published experiments, it was found that RP recordings were significantly enhanced when subjects performed the button press under the instruction to attend in particular to their intention to act, suggesting that the RP in fact reflects attention to volition or intention. Furthermore, one experiment compared the results of neurotypical subjects with those suffering from obsessive-compulsive disorder ('OCD'). Whilst neurotypical subjects displayed greater amplification in the later part of the RP when attending to their intentions (previously associated with conscious awareness of intention), subjects with OCD did not display this same increase. This suggests that OCD, like Tourette's Syndrome, likely involves abnormalities in brain activity associated with volition.

Finally, two novel experiments are of notable relevance. First, Houdayer, Lee and Hallett[19] conducted a relatively simple study where subjects were seated in an armchair for one hour whilst EEG activity was recorded, with the sole instruction not to close their eyes or fall asleep. RP recordings before spontaneous movements were subsequently compared with recordings for instructed movements taken during a separate session.

[18] Shiro Takashima, André M. Cravo, Koichi Sameshima and Renato T. Ramos, 'The effect of conscious intention to act on the bereitschaftspotential' (2018) 236(9) *Experimental Brain Research* 2287; Shiro Takashima, Fernando Araujo Najman and Renato T. Ramos, 'Disruption of volitional control in obsessive-compulsive disorder: Evidence from the bereitschaftspotential' (2019) 290 *Psychiatry Research: Neuroimaging* 30; Shiro Takashima, Carolina Y. Ogawa, Fernando Araujo Najman and Renato T. Ramos, 'The volition, the mode of movement selection and the readiness potential' (2020) 238(10) *Experimental Brain Research* 2113.

[19] Elise Houdayer, Sae-Jin Lee and Mark Hallett, 'Cerebral preparation of spontaneous movements: An EEG study' (2020) 131(11) *Clinical Neurophysiology* 2561.

Therefore, that RP recordings could be isolated for truly spontaneous movements which were not an artifact of experimental conditions. The RP was recorded in most subjects for movements made under both the spontaneous and instructed conditions. Spontaneous motions elicited the RP more strongly in medial frontocentral regions of the brain (including the pre-SMA, SMA, rostral cingulate zone and caudal cingulate zone) which the authors attribute potentially to the 'greater influence of internal triggering.'[20] Instructed motions elicited stronger RP amplitude in central regions potentially more reflective of the motor preparation of pre-planned specific movements.[21] The authors propose that the greater RP amplitude for instructed movements may reflect the greater attention that is being paid to those movements, again supporting the attention-to-volition hypothesis.

Second, Nann, Cohen, Deecke and Soekader[22] conducted a novel experiment taking EEG recordings from two subjects completing several bungee jumps from a 192-metre-high platform, comparing these with recordings taken from a jump height of only one metre. Again, the intention was to explore the RP in a potentially threatening, "real life" scenario without artifacts of experimental tasks. The researchers found high correlation in the RP waveforms recorded across the two conditions, suggesting that 'possible life-threatening decision making has no impact on the [RP's] spatiotemporal dynamics.'[23] These latter two experiments provide some modest demonstration of the RPs continuing involvement in motion and (attention to) volition outside of traditional laboratory conditions.

The experiments conducted most recently within the last five years have suggested that the RP may not necessarily reflect volition *per se*, but is that which brings conscious awareness to such underlying volition. This is especially hypothesised with regards to the late, steeper component of the RP, with the earlier slow RP being more associated with motor preparation. If correct, the implication follows that volition or intention occurs

[20] *Ibid*., 2563; citing Marie-Pierre Deiber, Manabu Honda, Vicente Ibañez, Norihiro Sadato and Mark Hallett, 'Mesial motor areas in self-initiated versus externally triggered movements explained with fMRI: Effect of movement type and rate' (1999) 81(6) *Journal of Neurophysiology* 3065.
[21] Houdayer, Lee and Hallett (2020), 2563 – 2564.
[22] Marius Nann, Leonardo G. Cohen, Lüder Deecke and Surjo R. Soekadar, 'To jump or not to jump – The bereitschaftspotential required to jump into 192-meter abyss' (2019) 9(1) *Scientific Reports* 2243.
[23] *Ibid*., 2247.

*prior* to the RP, with the RP instead reflecting that that intention (and coupled preparation to act) is being brought into conscious awareness. In other words, the brain unconsciously decides *when* to initiate voluntary action, whilst the individual becomes consciously aware of that pre-existing intention and associated motor preparation later, likely as a result of motor preparation itself and the (especially late) RP.

### *5.1.1. The Legal Relevance of the Initiation of Volitional Action*

This finding has important implications for the current approach to *mens rea* within the concept of legal responsibility. Introduced more fully in chapters six and eight of this thesis, the attribution of legal responsibility requires the coincidence of *mens rea* and *actus reus*; that is, the defendant must have possessed the requisite guilty state of mind (*mens rea*) at the time they committed a criminal act (*actus reus*). The quintessential textbook hypothetical concerns a defendant who resolves to shoot his enemy next week, thus forming the *mens rea* of an intention to kill, but then who accidently hits and kills his enemy whilst reversing his car out of the driveway. Whilst undoubtedly pleased with the happenstance outcome (and perhaps guilty of manslaughter / causing death by dangerous driving), the hypothetical defendant is not legally responsible for murder, because his intention to kill did not coincide with his actual actions. In other words, it was purely accidental that the defendant hit his enemy with his vehicle whilst, in that moment, he had no intention of hitting anybody with his car at all.

The coincidence of *mens rea* and *actus reus* is premised on the presumption that people possess conscious control over their decisions and actions such that, when they commit prohibited criminal acts with the requisite intention or recklessness *etc.,* they are morally blameworthy for having freely chosen such a criminal course of action or for otherwise having failed to exercise conscious control over their actions to conform with the law. However, the implications of the previous discussion suggest that, just as the brain decides *what* and *how* to do a particular thing as a result of automatically processing networks or mechanisms, so that brain may similarly decide automatically *when* to initiate a particular action and, crucially, prior to conscious awareness of having made a decision or initiated its associated action.

In this case, the aforementioned legal presumption becomes untenable; the coincidence of a given criminal intention with the initiation of its relevant criminal action may both arise automatically and without the presumed free conscious choice, nor potentially without the possibility for conscious control or intervention (*e.g.,* a conscious veto). This latter possibility of a veto more closely relates to the *whether* component of a decision discussed in chapter six, below, whilst the present chapter proceeds to consider the timing of conscious awareness of volitional action, and the implications this holds for the role of consciousness in decision-making and control over decisions and actions.

## *5.2. Consciousness and Timing*

### *5.2.1. The Half-Second Delay in Consciousness*

A seminal series of experiments were conducted and published by Libet *et. al.* in 1964,[24] 1967,[25] 1975,[26] 1979,[27] and 1992,[28] investigating the 'neurophysiological activities of the cerebral cortex which may be involved in the elaboration or mediation of conscious sensation',[29] – *i.e.*, the processes within the brain which result in a conscious awareness of stimuli. The former three publications considered the impact of different variations of electrical stimuli in generating a conscious sensation in response; the latter two publications expanded upon the former by examining the subjective backwards referral of timings for conscious sensory experiences.

[24] Benjamin Libet, W. Watson Alberts, Elwood W. Wright Jr, L. D. Delattre, Grant Levin and Bertram Feinstein, 'Production of threshold levels of conscious sensation by electrical stimulation of human somatosensory cortex' (1964) 27(4) *Journal of Neurophysiology* 546.
[25] Benjamin Libet, W. Watson Alberts, Elwood W. Wright Jr and Bertram Feinstein, 'Responses of human somatosensory cortex stimuli below threshold for conscious sensation' (1967) 158(3808) *Science* 1597.
[26] Benjamin Libet, W. Watson Alberts, Elwood W. Wright Jr, M. Lewis and Bertram. Feinstein, 'Cortical representation of evoked potentials relative to conscious sensory responses and of somatosensory qualities – in man' in Kornhuber H. H. (ed.), *The Somatosensory System* (Thieme 1975).
[27] Benjamin Libet, Elwood W. Wright Jr, Bertram Feinstein and Dennis K. Pearl, 'Subjective referral of the timing for a conscious sensory experience: A functional role for the somatosensory specific projection system in man' (1979) 102(1) *Brain: A Journal of Neurology* 193.
[28] Benjamin Libet, Elwood W. Wright Jr, Bertram Feinstein and Dennis K. Pearl, 'Retroactive enhancement of a skin sensation by a delayed cortical stimulus in man: Evidence for delay of a conscious sensory experience' (1992) 1(3) *Consciousness and Cognition* 367.
[29] Libet *et. al.* (1964), 546.

Experimental subjects underwent therapeutic surgical procedures that involved exposing the somatosensory cortex of the brain, which is associated with receiving and processing various sensory inputs such as touch, pain, temperature and proprioception (the position-in-space of different body parts). The procedures provided the opportunity for electrodes to be placed directly on the somatosensory cortex whilst the subjects remained conscious. Electrical pulses of varying intensity, polarity, duration and repetition frequency were applied whilst the awake subjects reported on their experiences, in particular on those occasions when electrical stimuli resulted in a conscious sensory experience. These experiences could manifest as 'natural-like' responses such as sensations of movement, pressure, vibration and temperature, or 'paraesthesia-like' responses such as tingling, electric shocks, pins and needles, and numbness.[30] Furthermore, the experiences reported from direct stimulation to the somatosensory cortex were compared with reports arising from peripheral skin stimulation achieved through electrodes applied directly to the skin.

The principle finding in these experiments was that 'substantial delays, of up to about 0.5 [seconds], before achieving cerebral "neuronal adequacy" appear to be required for eliciting a sensory experience.'[31] The 'most interesting and productive' relationship observed between the various electrical parameters related to the intensity of current applied and their train duration.[32] These were each adjusted to reach the same, 'just barely threshold' required to produce a conscious sensory experience for the subject, resulting in a 'minimum (liminal) intensity below which no sensation can be elicited' regardless of the length of the train duration. More pertinently, 'the liminal intensity stimulus elicits *no reportable sensory experience at all* unless its repetitive pulses are continued *for an average of 0.5* [*seconds*].'[33] From a neurological perspective, this is a remarkably long period of time before which conscious awareness of stimuli can arise, not least considering that the 'earliest neural messages [from sensory stimulation] reach the appropriate primary sensory cortex first, within 10-25ms.'[34] These findings apply equally to direct stimulation to the somatosensory cortex as they do to peripheral stimulation on

[30] Libet *et. al.* (1975), 300.
[31] Benjamin Libet, 'Brain stimulation in the study of neuronal functions for conscious sensory experience' (1982) 1(4) *Human Neurobiology* 235, 221.
[32] *Ibid*., 236.
[33] *Ibid*. (emphasis added).
[34] *Ibid*., 235.

the skin; *i.e.* the latter stimulation must still generate the requisite 0.5 seconds of neural activity within the brain before the conscious experience of peripheral stimulation can be perceived by the subject.

Libet *et. al.* further report how electrical pulses applied to the somatosensory cortex which did not meet the liminal intensity necessary to elicit a conscious sensory experience nevertheless resulted in 'substantial "direct cortical responses".'[35] Similarly, skin stimuli which were also below the threshold required to generate conscious sensation could 'still elicit a small primary evoked potential' within the brain such that, 'clearly, there can be substantial neuronal responses to stimuli in the sensory pathways that are not sufficient, and at least in some cases also not necessary, for eliciting conscious sensory experience.'[36] That notwithstanding, Libet suggests that it is probable that some such neural responses to sub-liminal stimuli 'could be involved in behavioral and psychological detection at unconscious levels.'[37] Stressing the separation which must be drawn between behavioural detection on the one hand and subjective experience of a stimulus on the other, Libet submits that this demonstrates how the 'former may be manifested with or without the latter', supporting the contention that subjective conscious experience is dependent upon specific kinds and durations of neural activity that are not essential for the unconscious detection of stimuli.[38]

A third important finding from the half-second delay in consciousness experiments concerns the phenomenon of a subjective retroactive referral of conscious sensory experiences. Although conscious awareness proceeds after a 0.5 second neural delay, this pause is not experienced in practice. Rather, 'there occurs an *automatic referral of the experience backwards in time*', with the primary evoked potential providing the timing signal for this retroactive referral.[39] The primary evoked potential refers to the first electrical signals which reach the somatosensory cortex between 10 – 25ms following sensory stimulation. Consequently, conscious awareness is experienced as if it were

[35] *Ibid*., 237.
[36] *Ibid*., 238.
[37] *Ibid*.
[38] *Ibid*.
[39] *Ibid*., 239 (original emphasis).

simultaneous with the stimuli which causes it, even though the conscious experience is actually produced after a half-second delay.

The publications from 1979 and 1992 explore this phenomenon in greater detail. Skin stimuli were paired with cortical stimuli at the requisite liminal intensity and duration to evoke conscious sensory experience, with the timing *between* each stimulus varying across trials. Where a cortical stimulus was applied after a skin stimulus following a delay of 300-400ms or more between the two, subjects reported 'no appreciable subjective delay for the skin-induced sensation relative to the delayed cortically induced sensation.'[40] Moreover, subjects even reported skin-induced sensations to have occurred first although the skin stimuli had been applied a 'few hundred [milliseconds] after the onset of the cortical train.'[41]

It was further demonstrated how a 'cortical stimulus can retroactively enhance the sensation of a preceding skin stimulus, even though [the cortical stimulus] train does not begin until 300 – 400ms or more after that skin stimulus.'[42] It was hypothesised that '(1) the early primary evoked neural response' – *i.e.* the initial signals from sensory receptors which first reach the somatosensory cortex within as little as 10 – 20ms following stimulation – 'acts as a timing signal, and (2) there is a subjective referral of the timing of the skin-induced experience, from its actually delayed appearance back to the time of the initial fast primary evoked response of the cortex.'[43] No similar antedating would occur in relation to a cortical stimulus alone, however, because this did not generate the early primary evoked neural response in the absence of receiving sensory signals from below.

The hypothesis was tested and confirmed through trials comparing the effects of skin stimuli with stimuli applied to the medial lemniscus – part of the subcortical pathway leading to the somatosensory cortex.[44] In contrast to direct stimulation to the cortex itself,

[40] Benjamin. Libet, 'Cerebral physiology of conscious experience: Experimental studies in human subjects' in Osaka N. (ed.), *Neural Basis of Consciousness* (John Benjamins Publishing 2003), 63.
[41] *Ibid*.
[42] Libet *et. al.* (1992), 372.
[43] Libet (2003), 63.
[44] See Libet *et. al.* (1979).

the medial lemniscus *does* produce the early primary evoked neural response; however, similarly to the somatosensory cortex, it also requires a train duration of up to 500ms. As the hypothesis predicted, stimulation to the medial lemniscus had no reported delay relative to the skin stimulation despite being induced from 200 to 500ms later. This subjective referral backwards in time not only corrects the temporal distortion caused by the half-second neural delay required to elicit conscious experience, but 'if the appearance of a sensory experience is delayed by 400ms or more, a delayed cortical stimulus could modify the content of the experience before the experience finally appears.'[45]

### *5.2.2. The Unconscious Cerebral Initiative*

The most famous series of experiments by Libet *et. al.* concern what he termed as the "unconscious cerebral initiative", investigating recordable cerebral activity in relation to free and voluntary actions and, in particular, the point in time at which such activity arose relative to the subjects' subjective conscious experience of intending to make that action. The key findings are published in 1982b,[46] 1983a[47] and 1983b.[48] The first publication discusses the readiness potential ('RP') as an indicator of neuronal activity preceding different types of self-paced motor action, with the study designed to minimise 'all external factors that might affect the immediate initiation of a freely voluntary motor act.'[49] The two publications from 1983 explore the potential relationship between these RPs and the initiation of free and voluntary acts, considering in particular the time at which they occur relative to conscious awareness of a decision to act.

Subjects observed a cathode ray oscilloscope with a spot of light circulating in revolutions of 2.56 seconds, which effectively acted as a clock. At a time of their choosing, the subject

[45] Libet *et. al.* (1992), 372.
[46] Benjamin Libet, Elwood W. Wright Jr and Curtis A. Gleason, 'Readiness-potentials preceding unrestricted "spontaneous" vs pre-planned voluntary acts' (1982) 54(3) *Electroencephalography and Clinical Neurophysiology* 322.
[47] Benjamin Libet, Elwood W. Wright Jr and Curtis A. Gleason, 'Preparation- or intention-to-act, in relation to pre-event potentials recorded at the vertex' (1983a) 56(4) *Electroencephalography and Clinical Neurophysiology* 367.
[48] Benjamin Libet, Elwood W. Wright Jr, Curtis A. Gleason and Dennis K. Pearl, 'Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act' (1983b) 106(3) *Brain: A Journal of Neurology* 623.
[49] Libet *et. al.* (1982), 322.

would make a small voluntary motor action such as flexing the wrist, the timing of which was recorded from muscular electrical activity using an electromyogram ('EMG'). The subjects were encouraged to make the movements spontaneously when they felt the "urge" to move, rather than planning ahead to move at particular times. After making such an action, the subjects were asked to recall the "clock" position when they first became aware of a subjective will or desire to move, providing the timing of the 'W judgment.' An electroencephalogram ('EEG') was used to record 'preparatory cerebral processes' through measuring the RP, 'a scalp-recorded slow negative potential shift that begins up to a second or more before a self-paced act… [and] can also precede self-initiated "freely" voluntary acts which are not only fully endogenous but even spontaneously capricious in origin.'[50]

The onset time of the RP was 'found to be consistently in advance of W, the time of initial awareness of wanting to move' across both the average values for all series of trials and for each individual series of self-initiated acts that provided a simultaneous recording of the RP and W.[51] Although the RPs for each event within a series needed to be averaged in order to produce the recorded RP, statistical and mathematical evaluation 'strongly supported the view that each individual RP precedes each conscious urge.'[52] Moreover, this timing relationship was maintained regardless of which parameters were preferred for measuring the onset of the RP or for timing W using either of two modes of recall available to the subjects. Libel *et. al.* separate the results into two classes – "type I" and "type II" RPs – which refer respectively to when the subject reported experiencing some degree of preplanning or preparation to act, and when they acted spontaneously and endogenously without any sensation of preparation. For type I RPs, a 'ramplike' RP was recorded with an onset occurring on average at –1050ms (±175) prior to W; for type II RPs, the RP onset occurred on average at –550ms (±150) preceding W.[53]

[50] Libet *et. al.* (1983b), 624.
[51] Benjamin Libet, 'Unconscious cerebral initiative and the role of conscious will in voluntary action' (1985) 8(4) *Behavioral and Brain Sciences* 529, 532 – 533.
[52] *Ibid*., 533.
[53] *Ibid*., 532.

Interestingly, the subjects in trials exhibiting type I RPs reported both some experience of preplanning or preparation, as well as a more specific urge or intention to act just before they performed each motion. Those subjects 'clearly distinguished this urge or intention from any advance feelings of preplanning to move within the next few seconds.'[54] Libet reports further experiments where a 'large ramplike RP' resembling the type I RP was recorded in subjects who had been instructed to pre-plan their action; he therefore concludes that 'the RP component that starts at about –550 [milliseconds], the one that predominates in type II RPs recorded… is the one uniquely associated with an exclusively endogenous volitional process.'[55] With the onset of the RP occurring some –550ms prior to W, this conscious awareness of an intention to act was found to occur an average of –200ms prior to the initiation of movement; 'that is, subjects reported becoming consciously aware of the urge to move 200ms before the activation of the muscle' as recorded by the EMG.[56] This leaves an average delay of around 350ms between the '"physical" (cerebral) process preceding the "mental" (conscious intention).'[57]

### *5.2.3. Libet's Interpretation*

Regarding the half-second delay in consciousness, Libet concludes that an average of 500 milliseconds of relevant integrated patterns of neuronal activity – termed as the 'state of "neuronal adequacy"' – must expire *before* a conscious experience of stimuli is achieved.[58] He posits that '[o]ne viable hypothesis suggests that it is sufficient durations *per se*, of appropriate neuronal activities, that gives rise to the emergent phenomenon of subjective experience.'[59] Stimuli which fail to reach neuronal adequacy can, and in some cases do, still exhibit neural and behavioural responses; however, these occur at the subconscious level and do not result in a conscious sensory experience. Libet highlights how meaningful responses to sensory stimuli requiring both cognitive and conative processing have been quantitatively measured in reaction time tests at as little as 100 – 200 milliseconds, and he raises regular anecdotal observations such as reacting whilst

[54] *Ibid*.
[55] *Ibid*.
[56] *Ibid*.
[57] *Ibid*., 532 – 533.
[58] Libet (1982), 238.
[59] *Ibid*.

driving a vehicle and sporting activities such as hitting as baseball at 90 miles per hour. He further cites Taylor and McCloskey[60] for providing direct empirical evidence that reaction times for visual signals were the same whether or not subjects reported a conscious awareness of the signal or were otherwise completely unaware. Libet continues:

> 'If actual conscious experience of the signal is neurally delayed by several hundred milliseconds, it follows that these quick behavioural responses are performed unconsciously, with no awareness of the precipitating signal, and that one may (or may not) become conscious of the signal only after the action.'[61]

More controversial are the findings that the onset of preparatory brain activity as expressed through the RP 'regularly begins at least several hundred [milliseconds] before reported times for awareness of any intention to act in the case of acts performed *ad lib*.'[62] From this, Libet concludes that the cerebral initiation of even an entirely spontaneous action 'can and *usually does begin unconsciously*.'[63] Whilst the voluntary acts being studied were appreciably trivial – flexing the wrist or pressing a button – he considers that they may nevertheless be regarded as 'paradigmatic examples of unrestricted action', writing that the 'basic initiating process for these simpler volitional acts may be the same as that for the actual motor expression of other, more complex forms of voluntary actions, since the latter are manifested behaviorally only when final decisions to move have been made.'[64]

As Libet notes, 'many, if not most, mental functions or events proceed without any reportable awareness', from the detection of, and behavioural responses to, sensory stimuli to the cerebral initiation of voluntary acts.[65] Furthermore, the significance of unconscious processing is demonstrable even in relation to 'complex functions', such as

[60] Janet L. Taylor and D. I. McCloskey, 'Triggering of preprogrammed movements as reactions to masked stimuli' (1990) 63(3) *Journal of Neurophysiology* 439.
[61] Libet (2003), 74.
[62] Libet (1985), 536.
[63] *Ibid*.
[64] *Ibid.*, 532.
[65] Libet (2003), 68

the processes engaged in deliberation, problem-solving and creative thought.[66] If the majority of mental processing for even complex deliberation or tasks is occurring without specific conscious awareness, this raises the significance of even the brief period of conscious awareness of an intention to act studied by Libet. There is some support for this view to be found in peripheral studies and replications, discussed further in section 5.2.4, below.

> 'Put another way, the brain "decides" to initiate or, at least, to prepare to initiate the act before there is any reportable subjective awareness that such a decision has taken place.'[67]

Naturally, Libet proceeds to ask what role, if any, consciousness plays in the decision-making process if freely voluntary actions are initiated unconsciously in the brain 'well before the person consciously knows he wants to act.'[68] The studies revealed that conscious awareness of an intention to act arose approximately 200ms prior to electrical activity signalling the beginning of motor action. This can be reduced by 50ms to correct for errors in the subjects' timing for their conscious awareness, with this correction being provided by control tests where the subjects were invited to report the timings of electrical stimuli applied to the skin using the same cathode ray oscilloscope method. Libet considers that this interval of 150ms between conscious awareness and motor action 'would allow enough time in which the conscious function might affect the final outcome of the volitional process.'[69] He clarifies, however, that this interval is only 100ms long, with the final 50ms prior to muscle activation being the approximate time necessary for the relevant electrical signals to proceed from the primary motor cortex and activate spinal motor nerve cells. During this latter time, the act continues through to completion without any further possibility of being halted by the cerebral cortex. Nevertheless, he suggests that this would allow enough time for the consciousness function to exert an effect.

---

[66] *Ibid.*
[67] Libet (1985), 536.
[68] Benjamin Libet, 'Do we have free will?' (1999) 6(8-9) *Journal of Consciousness Studies* 47, 51.
[69] *Ibid.*

One proposition is the "conscious veto" whereby the conscious will might stop or prohibit the final progression of a volitional act such that no actual muscle action is initiated. He writes, '[c]onscious-will could thus affect the outcome of the volitional process even though the latter was initiated by unconscious cerebral processes.'[70] Indeed, Libet *et. al.* conducted a series of trials where the subjects were instructed to perform the motor action (flexing the wrist) at specified pre-set "clock times" as opposed to at their own volition; the EMG could then compare the timing of their action with the actual pre-set when the action was instructed to occur.[71] Furthermore, these results were compared with another series of trials where the same subjects were instructed to veto their intention to act just prior to the instructed time.

This revealed that subjects could trigger the motor action to occur within 50 – 100ms of the instructed time; more crucially, they could veto the intention to act within approximately 100 – 200ms of the instructed time to act, with the corresponding RP reversing direction concurrently within 150 – 250ms of the instructed time.[72] Libet *et. al.* note in particular that the development of the RP could be observed even when the subject knew beforehand that they would veto the action, and no motor action was recorded by the EMG. They posit that this "covert RP" 'might be a general feature of non-consummated urges or intentions to act.'[73] Anecdotally, this could occur when an individual experiences an intention to perform some act with socially unacceptable consequences, and subsequently prevents themselves from acting at the last moment.

Libet also considers whether conscious will could serve as a final "trigger" to enable the volitional process to proceed through to action. However, he dismisses this suggestion for lacking any evidence in support, unlike the veto function. He suggests further that when voluntary acts become somewhat automated, they exhibit a comparatively minimal RP and an absence of reportable conscious intention to act; automatic acts can 'clearly go to completion without any conscious trigger available.'[74] In relation to the role of

[70] *Ibid.,* 51 – 52.
[71] See Libet *et. al.* (1983a).
[72] *Ibid.,* 369.
[73] *Ibid.*, 371.
[74] Libet (1999), 52.

consciousness as a veto function in decision-making, Libet asks whether this conscious veto could itself be the result of a preceding unconscious process, in the same manner as the development of a conscious will to act in the first place. As he writes, '[i]f the veto itself were to be initiated and developed unconsciously, the choice to veto would then become an unconscious choice of which we *become* conscious, rather than a consciously causal event.'[75] He concludes both that there is 'no logical imperative' requiring there to be specific neuronal activity that precedes a conscious *control* function, and nor is there experimental evidence precluding the possibility of such a control process appearing without prior unconscious processes.[76] However, as is discussed in the following section, it is likely that some of the unconscious processes leading to the conscious veto have since been discovered.

#### *5.2.4. Replications, Updates and Further Support*

The experiments by Libet *et. al.* were conducted more than 30 years ago in the 1980s and were restricted by the technologies available at that time. Two important modern variations have been conducted by Soon *et. al.* in 2008[77] and 2013;[78] the former considered voluntary motor action decisions similarly to Libet's investigations, whilst the latter went further to look at more complex and abstract decisions which did not involve motor action. Crucially, an updated method of timing the subjective awareness of reaching a conscious choice was employed, whilst brain activity was observed in considerably greater detail using functional magnetic resonance imaging ('fMRI'), both of which address key methodological criticisms surrounding Libet's original paradigm.

In each of the experiments cited, subjects were invited to make a choice between two options – the former experiment concerned a choice between pressing a left or right button, whilst the latter involved the choice between adding or subtracting two numbers and thus did not involve any related motor action. Subjects were invited to make a spontaneous

[75] *Ibid*.
[76] *Ibid.,* 53.
[77] Chun Siong Soon, Marcel Brass, Hans-Jochen Heinze and John-Dylan Haynes, 'Unconscious determinants of free decisions in the human brain' (2008) 11(5) *Nature Neuroscience* 543.
[78] Chun Siong Soon, Anna Hanxi He, Stefan Bode and John-Dylan Haynes, 'Predicting free choices for abstract intentions' (2013) 110(15) *Proceedings of the National Academy of Sciences* 6217.

free choice at a time of their choosing, and were to recall from a stream of letters appearing on a screen which letter had been present when they first became aware of having made a conscious choice. In the motor action study, subjects would then immediately select their choice between pressing a left or right button. In the abstract choice study, subjects would recall the letter appearing on the screen when they reached a conscious decision to perform addition or subtraction, and then would perform that operation on two simple numbers appearing on the next screen, finally selecting the answer on the third screen which randomly positioned the correct addition and subtraction answers along with two incorrect answers.

These experiments present a number of advantages over the original studies conducted by Libet *et. al.* First, as the RP is generated in the SMA of the brain it can only provide information regarding the later stages of motor planning, whereas the fMRI could observe in near real-time which different areas of the brain were being engaged in making a decision, *and* make these observations much earlier in the decision-making process than when the RP begins. In so doing, the use of fMRI helps to illuminate whether such decisions originate in the SMA where a decision to move is instigated, or whether further 'high-level planning stages might be involved in unconsciously preparing the decision.'[79] Second, with the time delay between the onset of the RP and conscious awareness of a decision being only a few hundred milliseconds, the updated timing method for conscious awareness made greater allowance for misjudgements in the timing of brain activity and subjective awareness, with each screen updating every 500ms. Third, by investigating choices between different options, the studies consider whether 'any leading brain activity… selectively predict[s] the specific outcome of a choice ahead of time.'[80]

In relation to voluntary motor actions, 'two brain regions encoded *with high accuracy* whether the subject was about to choose the left or right response *prior to the conscious decision*.'[81] Specifically, predictive fMRI signals from the frontopolar cortex (BA10) were present seven seconds prior to the subject becoming aware of making a conscious

[79] Soon *et. al*. (2008), 543
[80] *Ibid*.
[81] *Ibid*., 544 (emphasis added).

motor decision. Accounting for the delay in blood-oxygen-level dependent imaging ('BOLD') inherent within the method of using fMRI imaging, 'the predictive neural information will have preceded the conscious motor decision by up to [ten seconds].'[82] The second brain region where predictive signals were similarly recorded was in the parietal cortex, from the precuneus to the posterior cingulate cortex. Furthermore, the timing of a free motor decision could be predicted as early as five seconds prior to the reaching a conscious decision from both the pre-SMA and SMA, as well as from the frontopolar and parietal cortex 'just before' the decision.[83]

Whereas predictive RP signals were observed approximately –350ms prior to conscious awareness of a decision by Libet *et. al*., the predictive neural signals in the studies by Soon *et. al.* were observed significantly earlier, several seconds before conscious awareness emerged. The latter study further revealed a 'double dissociation' in the early stages of the decision-making process between those brain regions encoding the specific decision and those which determined the timing of the decision, whilst 'at later stages, right before the conscious decision, both of these regions begin to encode timing and handedness information.'[84] Although the timing mechanism used by Soon *et. al.* was inherently less sensitive than that by Libet *et. al.*, allowing greater room for potential misjudgements in the timing of subjective awareness of a conscious decision, subjects nevertheless became aware of an intention to act within –1,000ms prior to the motor action.[85] This accords with Libet's interpretation that conscious awareness of a decision to act only emerges in the very late stages prior to movement. Furthermore, the significant delay between predictive neural activity and conscious awareness of a decision to act more than accounts for any discrepancy in subjects' subjective timing of their conscious decision.[86]

Soon *et. al.* further investigated voluntary motor decisions where subjects' responses were instructed to be made at an externally determined time. Once again, the frontopolar

[82] *Ibid*.
[83] *Ibid*.
[84] *Ibid*.
[85] *Ibid*.
[86] *Ibid*., 545.

cortex was observed to provide predictive signals during the subjects' choice selection, with predictive information emerging in the precuneus after a selection had been made. As Soon *et. al.* write, one interpretation of these findings is that the 'frontopolar cortex was the first cortical stage at which the actual decision was made, whereas precuneus was involved in storage of the decision until it reached awareness.'[87] Taken together, they conclude that the frontal and parietal cortex contained 'considerable information' predicting the outcome of a motor decision that subjects 'had not yet consciously made', such that 'when the subject's decision reached awareness it had been influenced by unconscious brain activity for up to [ten seconds].'[88]

In the second series of experiments reported in 2013, Soon *et. al.* discovered that a medial frontopolar region and a region straddling the precuneus and posterior cingulate 'began to encode the outcome of the upcoming decision' up to four seconds prior to conscious awareness of that decision being made.[89] Predictive information for the timing of each decision was similarly recorded in different brain regions up to four seconds prior to the abstract decision, namely the pre-SMA, SMA and rostral cingulate zone, concurring with the 2008 study discussed above.[90] In the absence of a related motor action, the predictive information observed in the 2013 study cannot be attributed to the preparation of motor actions in the brain. Soon *et. al.* conclude that these results demonstrate how regions of the brain 'encode freely chosen abstract intentions before the decisions have been consciously made.'[91] Thus, although studying different brain activity to the RP which was investigated by Libet *et. al.*, these finding nevertheless support the ultimate conclusion that decisions are being formed unconsciously in the brain, with conscious awareness of those decisions – and thus, the possibility for conscious intervention – being limited to the very late stages of the decision-making process.

*

[87] *Ibid*.
[88] *Ibid*.
[89] Soon *et. al.* (2013), 6218.
[90] *Ibid.,* 6219.
[91] *Ibid.*

Fried *et. al.* have conducted two illuminating studies with implications regarding the role of "will" or intention to act in human behaviour, reported in 1991[92] and 2011.[93] In each study, subjects were undergoing therapeutic brain surgery for intractable seizures or epilepsy. The former study involved electrical stimulation mapping of the mesial frontal cortex, whereby electrical stimulation was applied directly to the brain whilst reactions were observed from awake patients. In the latter study, depth electrodes were implanted within the pre-SMA and SMA allowing for the recording of activity from individual neurons. The experimental paradigm by Libet *et. al.* was then performed, with patients pressing a button at a time of their choosing, whilst reporting the W time of each conscious urge to act; three subjects in the latter study performed a variation where they could select between pressing a button with their right or left index finger.

In the former 1991 experiment, the majority of responses to electrical stimulation were elicited in the form of overt bodily movements. Less frequently, however, stimulation resulted in reports of sensory experiences falling into three categories: sensations of tingling, numbness, warmth or pain; experiences of movement without any corresponding motor action; and experiences of an urge to move or anticipation of being about to do so.[94] For the latter two categories, Fried *et. al.* observed that 'these responses are often obtained only at threshold currents above which overt motor activity would be readily elicited.'[95] They conclude that the SMA is involved in the subjective experience of intention that accompanies motor activity, whilst Haggard interprets the study as suggesting that 'a conscious experience akin to intention is part of the normal neural preparation for voluntary movement.'[96]

The latter 2011 experiment is particularly interesting in providing the opportunity to observe in close detail the behaviour of individual neurons whilst the Libet *et. al.*

[92] Itzhak Fried, Amiram Katz, Gregory McCarthy, Kimberlee J. Sass, Peter Williamson, Susan S. Spencer and Dennis D. Spencer, 'Functional organization of human supplementary motor cortex studied by electrical stimulation' (1991) 11(11) *The Journal of Neuroscience* 3656.
[93] Itzhak Fried, Roy Mukamel and Gabriel Kreiman, 'Internally generated preactivation of single neurons in human medial frontal cortex predicts volition' (2011) 69(3) *Neuron* 548.
[94] Fried *et. al.* (1991), 3658.
[95] *Ibid.,* 3663.
[96] Patrick Haggard, 'Human volition: towards a neuroscience of will' (2008) 9(12) *Nature Reviews Neuroscience* 934, 942.

experimental paradigm was repeated. Fried *et. al.* found that 'preconscious activity of small assemblies of single neurones in the medial frontal lobe not only precedes volition but can also predict volition and its time of occurrence on a single trial basis.'[97] Furthermore, where potential inaccuracies in the subjects' reporting of W have been raised in relation to the original studies by Libet *et. al.*, Fried *et. al.* find that inaccuracies of up to –200 milliseconds did not significantly affect the number of neurons that changed their activity prior to W.[98] Further still, they observed that processes in the neurons studied began several hundred milliseconds, and sometimes several seconds before reports for W, and in the limited trials involving a choice between left and right this neural activity could be predictive of that decision. Fried *et. al.* conclude that these findings 'lend support to the view that the experience of will emerges *as the culmination of* premotor activity… starting several hundreds of [milliseconds] before awareness.'[99]

*

One potential issue with the studies presented thus far is the relative simplicity of the decisions being studied, even in the 2013 investigation of more complex abstract decision-making by Soon *et. al.* However, a study by Tusche *et. al.* in 2010[100] lends support to the submission that these results may be extrapolated to more complex decisions. The study considered the effect of attentional focus on a consumer decision by comparing the brain responses of two experimental groups recorded through fMRI imaging. The first "high attention" group were instructed to actively evaluate images of different cars and rate their attractiveness with fMRI recording brain responses in relation to these task-relevant images whilst they were the explicit focus of attention.

The second "low attention" group engaged in a visual fixation task whilst task-irrelevant images of cars were presented outside of the focus of attention. Control trials were further conducted with both groups later being presented with 20 images, of which ten had

[97] Fried *et. al.* (2011), 555.
[98] *Ibid.*
[99] *Ibid.*, 557.
[100] Anita Tusche, Stefan Bode and John-Dylan Haynes, 'Neural responses to unattended products predict later consumer choices' (2010) 30(23) *The Journal of Neuroscience* 8024.

featured in the experiment and ten were previously unseen. The difference in results between the two groups 'strongly suggest that attention was effectively removed from products in the low attention condition of [the] experiment.'[101] Subjects from each group were later instructed to imagine themselves in a consumer setting where they would decide upon the purchase of a new vehicle, and for each previously presented image were asked whether or not they would like to purchase the car. None of the subjects were aware that they would later be asked about their potential purchases in this manner.

In the high attention group, highly accurate predictive information was recorded across the prefrontal cortex ('PFC'), namely in the medial frontal gyrus, right dorsomedial PFC and the bilateral ventromedial PFC.[102] Furthermore, the left insula and parahippocampal gyrus 'were found to contain stable information about later product choices.'[103] For the low attention group, similarly accurate predictions of subsequent consumer choices were recorded from the left medial PFC and bilateral insula, with further predictive information being found in the left inferior parietal lobe and bilateral superior temporal gyrus. Crucially, the 'decoding accuracies in brain regions predicting subsequent consumer choices under high and low attention conditions *were found to be comparable*.'[104]

Tusche *et. al.* conclude that activation patterns across different brain regions are found to predict the choices studied, with a 'close match' of predictive regions independent of the spatial attention afforded to the products; *i.e.*, 'the amount of predictive information in these areas was comparably as high when task-irrelevant products were presented outside the focus of attention as when they were actively evaluated and attended to.'[105] Their 'key finding' is that a strong reduction in the degree of attention given to the various products considered 'does not affect the choice-predictive information.'[106] This supports the idea from Libet *et. al.* and, in particular, Soon *et. al.* that even complex and abstract (non-motor) decisions reach a high degree of completion *before* consciousness is engaged;

[101] *Ibid.,* 8027.
[102] *Ibid.*
[103] *Ibid.*
[104] *Ibid.,* (emphasis added).
[105] *Ibid.,* 8029.
[106] *Ibid.,* 8030.

‘complex and important economic choices can be prepared automatically, in the absence of explicit deliberation and without attention to products.’[107]

### *5.2.5. Summary Discussion on Consciousness and Timing*

The discussion in this section indicates towards Libet’s conclusions, that the timing of conscious awareness of reaching a decision appears to consistently fall later than the timing of the decision itself, which is first reached unconsciously. Indeed, this should not necessarily be surprising nor controversial from a neuroscientific perspective. It is reasonably trite to comment that, unless the conscious self exists as some *causa sui* homunculus in the brain, then all of conscious experience – including any such role that consciousness plays in decision-making itself – must be the result of prior activity in the brain which produces those conscious experiences. As Gazzaniga, Ivry and Mangun write, ‘[t]here is no question that we humans enjoy mental states arising from our underlying neuronal, cell-to-cell interactions.’[108] As the authors continue, however, ‘[t]hese mental states that emerge from our neuronal actions, such as belief, thoughts, and desires, in turn constrain the very brain activity that gave rise to them. Mental states can and do influence our decisions to act one way or another.’[109]

The criticism emerges, therefore, that Libet’s conclusions merely delay the point in time at which a decision is taken. That is to say, whereas a final decision to act may well indeed be initiated – and the requisite motor actions prepared – unconsciously before entering into conscious awareness as an urge or intention to act, we can nonetheless consciously contemplate and consider a decision before this point of action arises. In particular, this criticism suggests that Libet’s conclusions may well account for rapid decisions, but do not exclude a controlling role for consciousness in decisions that have been consciously deliberated over time. Afterall, conscious deliberation can “change our minds” with regards to a particular decision, in which respect it appears to have a causal and potentially controlling influence over those consciously deliberated decisions.

---

[107] *Ibid*., 8031.
[108] Gazzaniga, Ivry and Mangun (2019), 643.
[109] *Ibid*.

The particular contributions of conscious and unconscious thought to decision-making processes remains largely unknown, although it is considerably more likely that unconscious processes play a much greater role than those that enter conscious awareness. Nevertheless, this thesis posits one hypothesis flowing from the expanded model of decision-making presented in chapter two, above. Specifically, it is hypothesised that the conscious deliberation of any given decision affects that decision by affording greater time and mental resources for more different options to be considered in closer detail. That is to say, conscious deliberation of a decision can improve and, therefore, change a decision by devoting more time and energy to the underlying decision-making networks in the brain. Crucially, however, this does not entail that consciousness itself lends any greater degree of "control" over the outcome of a given decision. As all conscious experience results from prior unconscious neural activity, it would be a misnomer to describe consciousness as being capable of *controlling* a decision when any such conscious control as were possible would itself also be the result of prior unconscious activity. This would be the "truer" source of control although, of course, this unconscious activity was itself caused by prior factors and, *reductio ad infinitum*, the notion of direct control over the neural decision-making mechanisms in the brain ultimately evaporates.

Take the process of conscious deliberation itself, applying the model of decision-making presented in chapter two of this thesis and section 2.3 in particular. Competing neural networks across broad regions of the brain represent the various options under consideration for each of the *what, how, when, whether* and *why* components of a decision. However, our conscious experience is serial and, therefore, the content of each of these components can only be consciously experienced in turn. In deliberating *what* to get to drink, for example, we might consciously run through different options – water, coffee, beer – and attend to our feelings towards each option. In the process of evaluating those options, the competing neural networks representing each option are drawing from memories, emotions and the outcome of hypothetical future plans in order to attach value to each option, until the decision is finally reached according to whichever decision-strategy and stopping rules are being applied.

Critically, during conscious deliberation we experience the *outcome* of each of these underlying processes. We cannot, through the process of conscious deliberation, force a particular outcome from these processes. For example, in considering the decision of what drink to get, we might possess an old memory of disliking coffee which we do not immediately remember. It is possible to do things to try and improve or trigger that memory, such as by opening and smelling a bag of coffee; however, the fact of consciously deliberating the choice of which drink to get cannot *force* the retrieval of this particular memory. Nor does conscious deliberation force the competing neural networks to give more or less value to any available option. In our experience of deliberation, that memory will either appear, or it will not; we will experience a preference for one option or another. Equally, conscious deliberation does not *control* the competing networks representing the various *what, how, when, whether* and *why* components of a decision; rather, it lends greater time and mental resources to those networks, whilst the conscious experience remains that of becoming aware of the *outcome* of each component and the decision overall.

Applying the implications of the Libet experiments and their replications, therefore, even the conscious deliberation of any given decision is necessarily driven by underlying, unconscious cerebral processes over which there is no direct, conscious control. Every thought which enters consciousness during the deliberative process is itself the product of unconscious neural activity. In this regard, there is no problem of delaying the point in time at which a decision is taken; even a decision deliberated consciously over time is, inescapably, the product of prior unconscious neural activity. This same logic would apply to the decision to veto an action, within which Libet placed the last reserve of "free will" in the guise of the conscious veto. Naturally, this proposition is intimately connected to the *whether* component of a decision explored in chapter six below and, as is discussed there, evidence suggests that unconscious cerebral activity preceding Libet's conscious veto has, indeed, now been discovered.

Consider ***figure d*** in section 2.3.1 of this thesis, above, whereby the vertical line towards the left of the graph represents a short-time stopping rule for a decision taken quickly. At this point, the neural networks representing option B possess the greatest valence and this

becomes the decision. Now disregard the left vertical line and consider instead that the decision is pondered consciously and deliberated over time – option A becomes the decision when the neural networks representing this option reach a maximum threshold at which the decision is reached. What has the process of conscious deliberation contributed here? Most crucial is time; when the decision is taken earlier, option B possesses the greatest value, whereas when the decision is granted more time for consideration, option A becomes the preferred outcome. Second, it is likely that conscious deliberation assigns greater mental resources, energy or effort to the neuronal networks engaged in each component of the decision. This might explain why, for example, the value of option A exceeds that of option B over time, perhaps because certain memories associated with option A are older and require greater time or resources in order to be elicited. It is in these two regards that conscious deliberation can be considered as affecting the outcome of a decision; however, as has been discussed, it does not necessarily follow that conscious deliberation affords any degree of active, direct or online *control* over the outcome of that decision whilst, after all, it is this decisional outcome which potentially comprises the *mens rea* of a criminal offence.

*

Whilst the above discussion posits that consciousness may indirectly change but not *control* a decision deliberated over an extended period of time, it is pertinent to note that this particular hypothesis is not crucial for the purposes of discussing *legal* responsibility. In particular, the law requires that there is a coincidence of *actus reus* and *mens rea* in order to impart responsibility, which may be termed the "principal of coincidence."[110] The common analogy follows that, if a person one day determines to kill his rival (thus forming the requisite *mens rea* for murder), and the next day accidentally hits and kills his rival whilst reversing a car out of the driveway (thus forming the requisite *actus reus*), there is no coincidence between the *mens rea* and *actus reus*; the killing was accidental and without the requisite intention for legal responsibility.[111] The absolute rule is that the

[110] See further Michael Allen and Ian Edwards, *Criminal Law* (15th ed. Oxford University Press 2019), 66; Nicola Monagahan, *Criminal Law Directions* (6th ed. Oxford University Press 2020), 77 – 79.
[111] For examples in jurisprudence, see *R v Jakeman* (1983) 76 Cr App R 223.

requisite *mens rea* must exist at the time when the *actus reus* is committed, albeit one may begin to exist prior to the other. For example, in the classic case of *Fagan v Metropolitan Police Commissioner*,[112] the defendant accidentally drove their car onto a policeman's foot, thereby establishing the *actus reus* (but not the *mens rea*) for the offence of assaulting an officer during the execution of their duty. However, the defendant then intentionally left the car in place upon becoming aware that it was on the officer's foot, from which point on he possessed both the requisite *mens rea* whilst committing a continuous *actus reus*.[113]

A criminal act that is first deliberated over a period time will establish *mens rea* and *actus reus* in the opposite order to that in *Fagan*. That is to say, an individual may first deliberate and plan a particular criminal act such as murder, during which time they form the requisite *mens rea* of an intention to kill. At this stage, however, no criminal act has been committed; the law requires that some, at least minimal, form of criminal act is initiated – the *actus reus* – before any legal responsibility can attach thereto. Thus, the individual who has planned a murder must proceed to perform some minimal act before they could be charged and convicted with an offence – *i.e.,* they must at least make some attempt to carry out that intention to kill. Arguably the single, most minimal act that can potentially amount to a criminal offence is speech, which may be used to assault another (*i.e.,* to cause another to apprehend immediate and unlawful violence),[114] incite or encourage others to commit criminal acts,[115] or be used in such an offensive manner as amounts to harassment or hate speech.[116] Even such a minimal act as talking, however, must ordinarily be accompanied at the same time by the requisite *mens rea* following the principal of coincidence.

With this in mind, it may readily be argued that it is the final *whether* component of a decision which is the most critical to the question of legal responsibility. An individual may select the goal of killing another, deliberate and plan how to carry out the act *etc.*,

[112] *Fagan v Metropolitan Police Commissioner* [1969] 1 QB 439.
[113] See also *R v Thabo-Meli* [1954] 1 WLR 228; *R v Church* [1966] 1 QB 59; *R v Le Brun* [1991] 4 All ER 673.
[114] *R v Savage and Parmenter* [1991] 1 AC 699, 740; *R v Ireland* [1998] AC 147.
[115] Serious Crime Act 2007, ss. 44 – 46.
[116] Public Order Act 1986, Parts 3 & 3A; Crime and Disorder Act 1998, ss. 28 – 32; Criminal Justice Act 2003, ss. 145 – 146.

but it is the final decision of whether or not to put that plan into motion which immediately precedes the minimal act that is required to establish the *actus reus* in coincidence with the *mens rea*. There can be no legal responsibility unless and until this final whether decision proceeds to translate a criminal intention into a criminal act, even for crimes that have been deliberated and planned in detail (but, crucially, not yet initiated into any form of action). Indeed, it might (slightly facetiously) be argued that writers of fictional crime books, films and television productions often plan criminal acts in detail for the purposes of their work albeit, of course, with no actual intention of putting their fictional work into action.

Whatever the role of consciousness in decisions deliberated over time, it is the particular role played – or, as the case appears to be, *not* played – by consciousness in the final process of initiating a decision into action that is of most critical relevance to criminal liability specifically. Thus, after the initiation of volitional action – which, following section 5.1 of this chapter, appears to be triggered by unconscious processes and without the direct involvement of conscious choice or control – it is the *whether* component of a decision that is arguably of the greatest importance for the purposes of legal responsibility. Indeed, it is within this final component that Libet attempted to preserve some manner of conscious control over decisions by theorising the possibility for a conscious veto. This is explored in closer detail in the following chapter six of this thesis.

## *5.3. From Volition and Consciousness to Legal Responsibility*

As was set out briefly in section 2.4, above, the law assumes that all adults possess the capacity to grasp and be guided by good reason and to exercise conscious control over decisions and actions, unless the contrary is demonstrated. The evidence discussed in this chapter, however, calls into question the precise relationship between consciousness and volitional *control* over decisions and actions. With sensory signals reaching the somatosensory cortex in as little as 10 – 25 milliseconds and behavioural responses possibly arising in as little as 100 – 200 milliseconds thereafter, this can leave between 275 – 390 milliseconds after behaviourally responding to a stimulus before a subject may be consciously aware of either the stimulus or their unconsciously triggered response.

Whilst this particular finding may be most relevant to rapid, "snap" responses to a given set of circumstances, such impulsive decisions can nonetheless have significant practical and legal consequences, such as where an individual is accidentally shoved in a bar and unthinkingly responds by immediately throwing a punch. Following the evidence discussed in this chapter, it is possible for such rapid but nonetheless criminal responses to be triggered before an individual becomes fully consciously aware of how they are responding. Indeed, there can be few people in the world who have not at some time experienced acting quickly and impulsively to a situation before thinking about it fully and consciously.

Questions may also be asked regarding the role that consciousness plays in decisions deliberated consciously over a longer period of time. Libet's studies concerning the unconscious cerebral initiative and the various modern replications such as by Soon *et. al.* suggest that a freely endogenous volitional decision to act is initiated within the brain long prior (in neurological terms) to any conscious awareness of having reached a decision, with such conscious awareness only arriving within the final second before motor activity. If this is extrapolated backwards and applied to a decision consciously deliberated over time, it stands to reason that each thought, consideration and conclusion that is experienced consciously during the deliberative process is itself the product of prior unconscious cerebral activity, as is concluded from the Libet paradigm. That is to say, consciousness *per se* cannot act as a *causa sui* in the deliberative process; everything that enters into consciousness during a decision deliberated over time must be the product of prior unconscious cerebral activity.

If conscious *awareness* of a decision is required prior to that decision being reached in order for conscious *control* to be exerted over the decision, these neuroscientific studies appear to preclude the conscious control of volitional intentions to act. Libet attempts to preserve a role for conscious control over volitional actions in the guise of the "conscious veto" which he suggested may be possible in the final moments before the execution of a decision into motor action. However, there is no reason why a conscious veto would not itself, like any other conscious experience, be the product of prior unconscious cerebral activity. Therefore, it can be argued that any final conscious decision to veto an action is

equally the product of prior unconscious activity, in which case Libet's conscious veto would not provide any greater degree of conscious *control* over that decision to act or not. This is significantly the subject of the following section of this thesis discussing the *whether* component of a decision; however, it is pertinent to highlight here briefly that evidence for just such prior unconscious activity preceding a final decision to veto an action has been found.[117]

It is important to stress on the one hand that the experiments discussed do not remove any role altogether for consciousness in the longer-term deliberation of decisions which are not enacted until days, weeks or even years later. Clearly, we consciously contemplate thoughts, decisions and actions all of the time without proceeding each one through to physical activity. Moreover, there is well-documented empirical evidence on the role of consciousness in cognitive control processes such as response inhibition – preventing an otherwise intentional action, such as Libet's conscious veto; conflict resolution, which describes mechanisms of selection between competing response alternatives; and task-switching from one cognitive task to another.[118]

On the other hand, there is the strong suggestion that the general conclusions drawn from the Libet paradigm – whereby a conscious decision is caused by underlying cerebral activity occurring at an unconscious level – would be equally applicable to other conscious processes engaged in deliberative contemplation.[119] Indeed, as discussed above, Brass and Haggard have identified such underlying neural processes giving rise to conscious response inhibition.[120] More generally, a philosophical dualism between the mind and the body has been 'widely rejected within psychiatry, psychology, and neuroscience'[121] such that any conscious activity – including the conscious control over actions – is presumed to have preceding unconscious processes which cause that conscious activity to emerge. The contrary presumption would regard conscious

[117] Marcel Brass and Patrick Haggard, 'To do or not to do: The neural signature of self-control' (2007) 27(34) *Journal of Neuroscience* 9141.
[118] See further Simon van Gaal, Floris P. de Lange and Michael X. Cohen, 'The role of consciousness in cognitive control and decision making' (2012) 6(121) *Frontiers in Human Neuroscience* 1.
[119] For example, see Tusche *et. al.* (2010).
[120] See Brass and Haggard (2007).
[121] Dov Fox and Alex Stein, 'Dualism and doctrine' in Patterson D. and Pardo M. S. (eds.), *Philosophical Foundations of Law and Neuroscience* (Oxford University Press 2016), 108.

processes as a “first cause”, thus becoming the proverbial “ghost in the machine”. From this revised presumption regarding mind-brain duality, conscious awareness must arrive prior to the completion of a decision if it is to exert control, whereas the evidence discussed in this chapter strongly asserts that conscious awareness arrives *after* a decision has been reached unconsciously. Any further conscious decision to alter or change an earlier decision would, by extension, itself only arise *after* that altering decision had been reached unconsciously, and so on.

Furthermore, with regards to the question of legal responsibility for volitional acts, it is readily arguable that the possibility or otherwise of conscious control immediately prior to action is more relevant than the involvement of consciousness in general thought and contemplation that is further removed in time from a particular unlawful act. The law does not, as of yet, assign responsibility for mere thought alone. An individual may hope or intend the demise of another, but the law does not prescribe responsibility unless and until some action is initiated towards that intention. For example, if it is considered that speech is the “minimal” act that might be performed to transform mere thought into some form of criminal act, an individual who intends to murder another does not commit any crime until, for example, they have expressed that intention to the victim – thereby potentially committing a technical assault by placing the victim in fear of unlawful violence – or until they have discussed plans to commit that murder with another, amounting to a criminal conspiracy.

In each instance, legal responsibility does not arise until thought has been translated into criminal action, that action being speech in the examples above. Therefore, the crucial point for considering the question of conscious control over action is not whether the individual consciously deliberated a criminal act some time ago, but whether the individual could consciously control the fact that they performed the minimal action required to transform their mere thought into a criminal act. The experiments discussed in this chapter strongly suggest that, at the point at which a volitional (and potentially criminal) act such as speech is performed, that act has been prepared and initiated unconsciously, with conscious awareness (and thus the possibility of conscious control over that action) only arising within the very latest stages of the decision to act.

Furthermore, the experiments discussed in the following chapter will similarly suggest that any such possibility for conscious control over vetoing an action (*i.e.,* the *whether* component of a decision to act) is similarly preceded by unconscious activity in the brain.

Going further, if the late conscious veto over an action is indeed also the result of prior unconscious cerebral activity as suggested by Brass and Haggard, then in causal terms the selection and preparation of a criminal act would be the result of unconscious decisions, *and* the failure to veto that act before its performance would equally be the result of an absence of the necessary unconscious cerebral activity required to produce the veto. Crucially, each of the selection, planning and initiation of the criminal act itself, and the critical failure to veto its performance, would fall entirely outside of the individual's conscious control, and rather would be determined by the existence or absence of the relevant unconscious cerebral processes required to initiate or prevent a "free" and conscious volitional act. It follows that whether or not an individual is held legally responsible for their action depends upon a factor over which that individual has no conscious control, namely the prior unconscious cerebral processes initiating or vetoing their criminal decision.

The above arguments apply equally to the capacity to consciously control decisions that is presumed within the legal concept of volition, and the reliance upon proof of the existence of subjective states of mind within the broader concept of *mens rea*. In each case, the law may be criticised for drawing the dividing line between criminal liability and absolution of responsibility upon a point of happenstance. With regards to subjective mental states – and drawing from the conclusions of the present chapter and previous chapters three and four – it might be regarded as happenstance whether or not unconscious brain activity gives rise to a criminal intention or other subjective *mens rea*. Regarding the presumed capacity for consciously controlling decisions and actions – and drawing from section 5.1 of the present chapter – it might also be regarded as happenstance whether or not unconscious brain activity initiates the relevant motor actions to perform a criminal act, and further happenstance whether or not the initiation of that volitional action coincides with the related subjective *mens rea* arising consciously in the mind. Finally, if the implications of the previous section 5.2 are taken to their conclusion – *i.e.,*

regarding the ability to control or veto the ultimate expression of a criminal decision through physical action, explored more fully below in chapter six – then it is happenstance whether or not unconscious brain activity actually arises in order to veto a criminal decision.

# 6. The *Whether* Component, Veto, and Impulse Control

> '[W]e humans think we are making all our decisions to act consciously and wilfully. We all feel we are wonderfully unified, coherent mental machines and that our underlying brain structure must somehow reflect this overpowering sense we all possess. It doesn't. Again, no central command center keeps all other brain systems hopping to the instructions of a five-star general. The brain has *millions* of local processors making important decisions. It is a highly specialized system with critical networks distributed throughout the 1,300 grams of tissue. There is no one boss in the brain. You are certainly not the boss of the brain. Have you ever succeeded in telling your brain to shut up already and go to sleep?'
>
> - Michael Gazzaniga, 2012.[1]

From the perspective of the law, the *whether* component of any decision to commit a criminal act is arguably the most important step. A person may form a decision to commit a particular criminal action (*what*), in a particular manner (*how*), at a given time (*when*), and with their reasons for so doing (*why*). But until some minimal step is taken to transform all of that intentionality into a physical action which comprises the *actus reus* of an offence – even if all that step amounts to is discussing the plan with another for the purpose of criminal conspiracy, or taking the initial steps to comprise a criminal attempt – no criminal offence has been committed. That is to say, it is not *unless and until* the *whether* component of a decision triggers criminal action that a person may be held legally responsible for a crime. This reflects the principle that criminal censure does not attach to mere thoughts, but only actions (and, less frequently, omissions). In this regard, a person can only ever be responsible for having *done* (or not *done*) some act.

[1] Michael S. Gazzaniga, *Who's in Charge?: Free Will and the Science of the Brain* (Robinson 2012), 43 – 44.

And, indeed, this arrangement accords with our daily experiences; virtually all people have likely fantasised at one time about doing something that could be criminal, whether shoplifting a tempting item from a store, exacting some revenge upon an enemy, or being dishonest on a tax return. Some people, such as novelists or screenwriters, might even devise fully-formulated criminal plans to provide a narrative to their work. But in each of these instances, whatever the degree of mental preparation and planning, and whether entertaining an actual intention to commit future criminal acts or purely fantasising, no offence can be committed until the decision of *whether* to act has been taken.

Furthermore, and again from a legal perspective, it is the *whether* decision that immediately precedes the actual committing of the *actus reus* of an offence that is of critical importance. An individual might decide in April that they are definitely going to rob a store in May, from which perspective the decision whether or not to act has, in one sense, already been taken. However, the critical moment of criminality is that point in time when a person actually commits some minimal criminal act in conjunction with the requisite criminal mindset – the moment of coincidence of *actus reus* and *mens rea,* discussed further throughout chapters five and eight of this thesis. Thus, in the moment before the minimal criminal act is done, the individual must make a final choice as to *whether* or not to commit that minimal act; in this moment, the individual still has an opportunity to walk away without having committed any offence.

Appreciating the aforementioned point is crucial to focusing the scientific investigation of the *whether* component of decision-making, as the topic of self-control or self-regulation is vastly diverse. Indeed, the capacity for self-control has been argued to be one of the defining features of the human species, alongside significant intelligence in comparison to other species. Forgas, Baumeister and Tice write:

> ‘The capacity to forego immediate pleasures and resist current impulses to secure greater but delayed rewards is a hallmark of the pursuit of enlightened self-interest… it is [] abundantly clear that the capacity to delay gratification is vitally important to human well-being. Agriculture, for example would be impossible without delaying gratification, as would

> saving money or food for the future. Likewise, the long education process that enables people to achieve great things within and for human culture depends on the capacity to delay gratification.'[2]

In the above respects, self-regulation is about maintaining and pursuing long-term goals and delaying gratification in order to meet those goals. The *whether* decision that is of particular interest to the question of legal responsibility, however, is that which is in operation immediately prior to a discrete criminal act, which therefore provides the principal focus for research considered in this chapter. Self-control in this regard might more accurately be described as concerning impulse control and the ability to veto decisions and actions; albeit, it is not claimed that there is necessarily any clear dividing line between self-regulation of immediate day-to-day actions and long-term goals. After all, distant objectives can only even be achieved through a succession of immediate actions, even if those actions actually consist of resisting an impulse to do some positive act which would jeopardise that long-term goal. It is therefore plain to appreciate how self-regulation – consisting *inter alia* of setting and pursuing long-term goals, controlling impulses and vetoing actions – is implicated across a huge range of domains.

> 'In fact, most major social and personal problems that afflict people in modern, Western culture have some degree of self-regulation failure as a core part of the problem. Inadequate or misguided self-regulation is involved in drug and alcohol addiction, eating disorders, obesity, crime and violence, prejudice and stereotyping, cigarette smoking, underachievement at school and work, unwanted pregnancy, sexually transmitted diseases, debt, failure to save money, gambling, domestic abuse, and many more.'[3]

There exists clear evidence linking poor capacities for self-regulation with a significant range of social and personal problems encountered in modern Western societies. So too

---

[2] Joseph P. Forgas, Roy F. Baumeister and Dianne M. Tice, 'The psychology of self-regulation: An introductory review' in Forgas J. P., Baumeister R. F. and Tice D. M. (eds.), *Psychology of Self-Regulation: Cognitive, Affective, and Motivational Processes* (Psychology Press 2009), 6.

[3] *Ibid*., 5; see further Roy F. Baumeister, Todd F. Heatherton and Dianne M. Tice, *Losing Control: How and Why People Fail at Self-Regulation* (Academic Press 1994).

exists a growing body of evidence linking such social and personal problems as inequality and poverty, and low levels of parental / maternal education to elevated stress physiology in infancy which, in turn, correlates with poor executive function and lower intelligence and cognitive abilities throughout childhood and into adulthood.[4] For example, Blair *et. al.* have conducted a number of studies revealing the influence of adverse rearing environments on infant attention, emotion and stress, finding in particular that factors overrepresented in poverty – such as lower income, lower levels of maternal education, and reduced prototypically responsive maternal caregiving behaviour – are associated with elevated levels of the stress hormone cortisol in infants at 7, 15 and 24 months old.[5]

Remarkably, African American children were found to have higher levels of cortisol than their Caucasian counterparts even when controlling for family characteristics and parenting behaviour, whilst elevated cortisol, conditions of poverty and maternal caregiving 'fully explained observed associations between African American ethnicity with low executive function and IQ.'[6] Blair and Ursache interpret these results by reference to the considerably worse conditions of poverty that African American subjects in the study experienced relative to their Caucasian counterparts, in terms of income, maternal education, household crowding and neighbourhood safety. They write:

> '[I]t is likely that African American ethnicity in this sample represents a marker of deep and persistent poverty. As such, results suggest that noted racial gaps in cognitive ability and school achievement in the United States reflect, in addition to well-documented inequalities in educational

[4] See further Clancy Blair, 'Stress and the development of self-regulation in context' (2010) 4(3) *Child Development Perspectives* 181; Clancy Blair, 'Developmental science and executive function' (2016) 25(1) *Current Directions in Psychological Science* 3.

[5] Clancy Blair, Douglas A. Granger, Katie T. Kivlinghan, Roger Mills-Koonce, Michael Willoughby, Mark T. Greenberg, Leah C. Hibel, Christine K. Fortunato and Family Life Project Investigators, 'Maternal and child contributions to cortisol response to emotional arousal in young children from low-income, rural communities' (2008) 44(4) *Developmental Psychology* 1095; Clancy Blair, Douglas A. Granger, Michael Willoughby, Roger Mills-Koonce, Martha Cox, Mark T. Greenberg, Katie T. Kivlinghan, Christine K. Fortunato and Family Life Project Investigators, 'Salivary cortisol mediates effects of poverty and parenting on executive functions in early childhood' (2011) 82(6) *Child Development* 1970.

[6] Clancy Blair and Alexandra Ursache, 'A bidirectional model of executive functions and self-regulation' in Vohs K. D. and Baumeister R. F. (eds.), *Handbook of Self-Regulation: Research, Theory, and Applications* (2nd ed. The Guildford Press 2011), 310.

opportunity, *the adverse effects of poverty on stress physiology, with cascading effects on self-regulation and executive functions.'*[7]

Taken together, two broad points may be drawn from the aforementioned research. First, a vicious cycle exists between poor self-regulation and a number of factors such as poverty, lower income and lower educational achievement. On the one hand, poor self-regulation has been shown to negatively impact such factors as cognitive ability and academic achievement, attaining and retaining employment and managing finances, which are all conditions associated with poverty. On the other hand, these same factors are revealed to demonstrably impact upon even a young infant's stress physiology, which results in poorer executive functions and self-regulation into childhood and adulthood. Thus, a vicious cycle between conditions of poverty and poor self-regulation emerges.

Second, and having particular regard to the links between poor self-regulation and criminality, it is submitted that any system of legal responsibility must have regard to the fact that a great number of the causes of poor self-regulation can be attributed to failures in the society which that system of legal responsibility seeks to govern. That is to say, if it is plainly recognised that societal failures resulting in children growing up in poverty is a demonstrable cause of poorer self-regulation and, consequently, greater instances of crime, then it is incumbent upon a system of law to attempt to make some redress to those causes of criminality when dealing with the individual offender. It is envisaged that this would predominantly take place through a leaning towards rehabilitative theories of punishment, discussed further in chapter twelve, below.

## *6.1. The Marshmallow Test*

The seminal Stanford marshmallow experiment provides a foundational study of self-control. The lead researcher, Walter Mischel, describes witnessing three daughters close in age grow, from "gurgling" babies to "enchanting" toddlers, to conversational children who, after only a few years, could sit and wait for things that they wanted. He writes, 'I wanted to understand willpower, and specifically delay of gratification for the sake of

[7] *Ibid* (emphasis added).

future consequences,'[8] and he wanted to explore how this skill first develops in children and what strategies they might successfully employ. The paradigm is relatively simple; subjects first learn to trust the researcher through the use of a bell which, whenever rung by the subject, results in the researcher entering the room. The subject has a choice of toys or sweet rewards in the original paradigm whilst marshmallows became famous from later video recordings; but the subject is always given the option of having one cheaper reward immediately, or having two (or better and more expensive) rewards later if they wait for the researcher to leave and then return to the room. The subjects are probed to ensure that they understand the instructions and the increased reward available, then the researcher leaves the room and the experiment begins.

Mischel and colleagues conducted the paradigm in a range of contexts across the 1960s, both in the US and across other cultures, and with ages ranging from kindergarten to adulthood. The youngest subjects were aged between 3 years and six months to 5 years and 8 months and were recruited from a Nursery School attached to Stanford University in the US. It was initially predicted that, in experimental conditions where the delayed reward remained present and available to the child, subjects would increase their voluntary delay time due to enhancing their attention to the reward. Conversely, 'it was anticipated that the condition in which the child was left without either reward would make it most difficult to bridge the delay time and therefore lead to the shortest waiting.'[9] Contrary to predictions, however, children were able to wait the longest when the reward did not remain in the room with them. Those children were also the most successful at devising a range of simple but effective self-distraction strategies 'through which they spent their time psychologically doing something (almost anything) other than waiting.'[10] Amongst such strategies:

> '[I]nstead of focusing prolonged attention on the objects for which they were waiting, they avoided looking at them. Some children covered their

[8] Walter Mischel, *The Marshmallow Test: Understanding Self-Control and How To Master It* (Transworld Publishers 2014), 16.
[9] Walter Mischel and Ebbe B. Ebbesen, 'Attention in delay of gratification' (1970) 16(2) *Journal of Personality and Social Psychology* 329, 331.
[10] *Ibid*., 335.

> eyes with their hands, rested their heads on their arms, and found other similar techniques for averting their eyes from the reward objects. Many seemed to try to reduce the frustration of delay of reward by generating their own diversions: they talked to themselves, sang, invented games with their hands and feet, and even tried to fall asleep while waiting – as one child successfully did.'[11]

Mischel *et. al.* conducted a number of repetitions and variations of the paradigm in different contexts in order to elucidate factors that might contribute to subjects being better or worse at delaying gratification. For example, an early completion of the experiment in Trinidad in 1958 found differing performance in children aged 7 to 9 years depending upon their cultural and ethnic backgrounds, with a tendency towards immediate gratification for children growing up in fatherless households, and a tendency towards delayed gratification as the age of the children increased.[12] A further replication amongst subjects aged between 12 and 14 years contrasted performance between children in general education and in a reform school for juvenile delinquents.[13] No statistically significant differences were found according to age but, as predicted, 'a significantly larger proportion of delinquent subjects cho[se] immediate, smaller reinforcement.'[14]

A further study of teenagers in general education in Trinidad also used simple personality measures for traits of achievement, acquiescence and social responsibility, comparing these traits for subjects who consistently chose immediate reinforcement, consistently chose delayed reinforcement, or were inconsistent in their choices.[15] A similar pattern followed whereby the subjects who consistently delayed gratification scored significantly higher on measures for achievement and social responsibility, and moderately lower for

[11] *Ibid*; see further Walter Mischel, Ebbe B. Ebbesen and A. Raskoff Zeiss, 'Cognitive and attentional mechanisms in delay of gratification' (1972) 21(2) *Journal of Personality and Social Psychology* 204.
[12] Walter Mischel, 'Preference for delayed reinforcement: An experimental study of a cultural observation' (1958) 56(1) *Journal of Abnormal and Social Psychology* 57.
[13] Walter Mischel, 'Preference for delayed reinforcement and social responsibility' (1961) 62(1) *Journal of Abnormal and Social Psychology* 1.
[14] *Ibid*., 4.
[15] Walter Mischel, 'Delay of gratification, need for achievement, and acquiescence in another culture' (1961) 62(3) *Journal of Abnormal and Social Psychology* 543.

acquiescence traits.[16] Dozens of similar studies were conducted by Mischel and colleagues across the 1960s and 1970s and, allowing for the fact that subjects naturally performed across a spectrum, two broad stereotypes could be described at either end of that spectrum from the overall research.[17]

The "puritan character" predominantly chose larger, delayed rewards and was 'more likely to be oriented toward the future.'[18] This character tended to have higher scores for measures of "ego-control", achievement motivation, social responsibility and trust, and tended to be more bright, mature, aspirational, and with greater self-control over impulsivity. Furthermore, the puritan character was associated most often with middle and upper socioeconomic groups, and was related to a 'relatively high level of competence, as revealed by higher intelligence, more mature cognitive development, and a greater capacity for sustained attention.'[19] At the opposite end of the spectrum was the more "impulsive character". This stereotype predominantly preferred immediate over delayed gratification and was less likely to wait or work for larger, delayed gains. Associated with this character was a greater concern for immediate over future rewards and higher impulsivity, whilst this character was more strongly correlated with lower socioeconomic groups and with 'membership in cultures in which achievement orientation is low, and with indices of lesser social and cognitive competence.'[20]

### *6.1.1. Follow-up Studies*

The original studies by Mischel and colleagues revealed on the one hand the range of responses to delayed gratification tasks amongst subjects from a range of ages and backgrounds, from those who consistently delayed gratification for greater rewards to

[16] *Ibid*., 546 – 550.

[17] For a comprehensive overview, see Walter Mischel, 'Processes in delay of gratification' (1974) 7 *Advances in Experimental Social Psychology* 249; Walter Mischel, 'Theory and research on the antecedents of self-imposed delay of reward' in Maher B. A. (ed.), *Progress in Experimental Personality Research: Vol 3* (Academic Press 1966).

[18] Mischel (1974), 253; citing Stephen L. Klineberg, 'Future time perspective and the preference for delayed reward' (1968) 8(3) *Journal of Personality and Social Psychology* 253.

[19] Mischel (1974), 253 – 254; citing Paul F. Grim, Lawrence Kohlberg and Sheldon H. White, 'Some relations between conscience and attentional processes' (1968) 8(3) *Journal of Personality and Social Psychology* 239.

[20] Mischel (1974), 254.

those who consistently preferred immediate gratification, with others performing inconsistently between the two. On the other hand, Mischel's studies revealed hints of correlations between the two stereotyped puritan and impulsive characters, both relating to personality traits such as intelligence, maturity and motivation towards achievement, and relating to broader backgrounds and cultures such as determined by nationality and socioeconomic upbringing. So much is, perhaps, not terribly surprising given the great range of behaviours demonstrated in the human species. In addition, Mischel's original work elucidated some of the strategies that young children would develop when attempting to delay gratification in the presence of an immediately tempting reward.

It is the follow-up studies conducted by Mischel and others that are particularly illuminating, however. In the first such follow-up, Mischel, Shoda and Peake obtained personality ratings from the parents of some 59 teenage subjects who participated as children in the original studies some ten years prior.[21] A simple questionnaire intended to avoid excessive demands on time asked parents to rate how their child compared to his or her peers on academic, social, frequency-of-problems and coping measures, rating from 1 to 7 with the higher score representing stronger performance than peers. A second, longer measure consisted of a modified California Q-Set including 100 personality-relevant items to be sorted according to their descriptiveness of the subject.[22] Summarising the main results, those subjects who had been able to delay gratification for longer as children:

> '[A]re more verbally fluent; use and respond to reason; are attentive and able to concentrate; are planful and think ahead; are competent and skilful; are resourceful in initiating activities; are self-reliant and confident; become strongly involved in what they do; can be trusted and are dependable; are self-assertive; are curious, exploring, and eager to learn; and show concern for moral issues. These children also do not tend to go to pieces under stress or become rattled and disorganized; are less likely to

[21] Walter Mischel, Yuichi Shoda and Philip K. Peake, 'The nature of adolescent competencies predicted by preschool delay of gratification' (1988) 54(4) *Journal of Personality and Social Psychology* 687.
[22] *Ibid*., 689 – 690.

> appear unworthy or think of self as bad; are not shy and reserved or slow to make social contacts; are not stubborn; do not tease other children; do not revert to more immature behavior under stress; are not afraid of being deprived or concerned about getting enough; do not tend to be suspicious and distrustful; do not show mannerisms or rituals; are not unable to delay gratification or wait for satisfaction; are not jealous or envious; do not become rigidly repetitive or immobilized under stress; and do not withdraw or disengage when under stress.'[23]

Plainly, there are issues with how precisely many of these attributes could have been accurately measured, and it is crucial to recall that the results were based entirely on parents reporting their subjective comparisons of their own children to peers and, as such, subjectivity abound in the measurements taken. That being said, the data pointed towards a strong correlation between the time that subjects were able to delay gratification as young children and their cognitive, social and coping competences as adolescents.[24] The authors suggest that the ability to delay gratification for larger goals may play an increasingly pervasive and influential role throughout a child's maturation, and 'therefore becomes increasingly linked with indicators of adaptive coping.'[25]

A second follow-up study by Shoda, Mischel and Peake attempted to corroborate previous findings with a larger sample of subjects, and take advantage of that same opportunity to identify the 'particular psychological conditions in which children's delay of gratification behavior is more likely to predict relevant individual differences in developmental outcomes.' [26] Similar measurement methods were utilised as in the previous study, above; however the 1990 follow-up also included further data in the form of the subjects' SAT verbal and quantitative scores obtained in school, providing an objective measure of their educational and cognitive performance. Similar findings were made regarding the subjects' personality traits as compared to their peers and reported by

[23] *Ibid*., 690 – 691.
[24] *Ibid*., 692.
[25] *Ibid*., 694.
[26] Yuichi Shoda, Walter Mischel and Philip K. Peake, 'Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions' (1990) 26(6) *Developmental Psychology* 978, 978.

their parents; those subjects who had been able to delay gratification for longer in preschool were 'rated as more likely to exhibit self-control in frustrating situations, less likely to yield to temptation, more intelligent, and less distractable when trying to concentrate.' [27] Moreover, whilst the sample sizes of available SAT results were admittedly 'barely sufficient', the data nevertheless provided similar correlations between delay time in preschool and verbal and quantitative SAT scores.[28]

The follow-up study by Shoda, Mischel and Peake also revealed an interesting link to the effects of subjects having been shown a particular strategy for resisting temptation during the marshmallow test, as contrasted with those subjects who developed such strategies spontaneously. Specifically, whereas delay-time was significantly predictive of future performance for subjects who developed cognitive strategies spontaneously, it was neither strongly nor consistently predictive when subjects had explicitly been given cognitive strategies to use by the researcher.[29] The authors are appreciably cautious in noting that, despite the strength of their results and the corroboration of subjective measures with more objective data, it is important that correlations with the SAT scores accounted for only about 25% of the variance; this is undoubtedly significant, but not necessarily overwhelming. Nonetheless, they submit that the observed correlations could reveal that the 'qualities that underlie effective self-imposed delay in preschool may be crucial ingredients of an expanded construct of "intelligent social behavior" that encompasses social as well as intellectual knowledge, coping, and problem-solving competencies.'[30] The fact that such a wide variety of outcomes could be predicted from preschool delay-of-gratification times, alongside the apparent significance of developing cognitive strategies spontaneously rather than by demonstration, suggests that the capacity for self-control may be something that is either comprised of a strongly innate component or is otherwise moderately determined from a young age.

---

[27] *Ibid*., 982.
[28] *Ibid*.
[29] *Ibid*., 983 – 984.
[30] *Ibid*., 985; citing Ann L. Brown and Judy S. DeLoache, 'Skills, plans, and self-regulation' in Siegler R. (ed.), *Children's Thinking: What Develops?* (Lawrence Erlbaum Associates 1978); Nancy Cantor and John F. Kihltrom, *Personality and Social Intelligence* (Prentice-Hall 1987).

Ten years later still, Ayduk *et. al.* (in a research team including Mischel and Peake) conducted a further follow-up study, exploring in particular any links between the subjects' delayed gratification capabilities during preschool and their success in coping with rejection sensitivity in adulthood.[31] This followed research proposing that the effective regulation of negative arousal may be important not only for inhibiting undesired and impulsive behaviours that are induced by stress, but 'also may facilitate execution of problem-solving strategies.'[32] The authors explore this proposed link through a follow-up study on participants of the original delayed gratification paradigms, hypothesising that those subjects able to delay for longer in preschool would be better insulated in adulthood against negative interpersonal and personal consequences arising from anxious rejection expectations. Questionnaires were sent to both the adult subjects of the original paradigms and their parents, including items from the Rosenberg Self-Esteem Questionnaire,[33] from Hazan and Shaver's Adult Attachment Styles Questionnaire,[34] and a modified version of the California Child Q-Set.

In concurrence with previous follow-up studies, Ayduk *et. al.* found a significant interaction between subjects' delay of gratification during the preschool paradigm and rejection sensitivity ratings as extracted from both subjects' and their parents' questionnaire responses.[35] With regards to behavioural outcomes, subjects who delayed for longer as children had also attained higher levels of education by adulthood. Of particular note, however, those with high rejection sensitivity attained similarly high levels of education when they also delayed for longer as children, as compared against those with high rejection sensitivity but a low delay time.[36] This suggests that the capacity for delayed gratification in childhood may provide an insulating effect against rejection sensitivity during maturation and adulthood. In a similar vein, delayed gratification in childhood predicted a lower instance of hard drug used (cocaine / crack cocaine) in adult

[31] Ozlem Ayduk, Rodolfo Mendoza-Denton, Walter Mischel, Geraldine Downey, Philip K. Peake and Monica Rodriguez, 'Regulating the interpersonal self: Strategic self-regulation for coping with rejection sensitivity' (2000) 79(5) *Journal of Personality and Social Psychology* 776.
[32] *Ibid*., 776.
[33] Morris Rosenberg, *Conceiving the Self* (Basic Books 1979).
[34] Cindy Hazan and Phillip Shaver, 'Romantic love conceptualized as an attachment process' (1987) 52(3) *Journal of Personality and Social Psychology* 511.
[35] Ayduk *et. al.* (2000), 781.
[36] *Ibid*., 782 – 783.

subjects, even with high levels of rejection sensitivity. Further still, consistent insulating effects of the delayed gratification ability were demonstrated in teenagers in a second study which showed that rejection sensitivity was 'negatively related to self-worth and interpersonal functioning in [high-rejection sensitive] children *unless they had high [delayed gratification] ability*.'[37]

In a further follow-up study – now more than 30 years following the original paradigms – Schlam *et. al.* considered whether the ability to self-regulate in childhood might also be predictive of body mass index in adulthood.[38] The potential links between being able to delay gratification and body mass are perhaps readily discernible, whilst a growing body of research has similarly drawn links between self-control and weight gain in children and adolescents.[39] Allowing for differences accounted for by sex, the authors found that the ability to delay gratification in childhood was linked to a 'significant portion of variance (4%) in the composite measure of BMI… each additional minute that a child delayed gratification predicted a 0.2-point reduction in BMI in adulthood.'[40] As they note, whilst this result is not particularly large, the presence of such a statistically significant effect *more than three decades following the original marshmallow experiment* is highly noteworthy, and further suggests towards the wide range of faculties and traits that may be impacted by good or poor self-regulation developed during early childhood.

Finally, with regards to the original marshmallow test subjects, a 2011 study by Casey *et. al.*[41] sought to investigate first whether there existed correlations between subjects'

[37] *Ibid*., 786 – 787 (emphasis added).
[38] Tanya R. Schlam, Nicole L. Wilson, Yuichi Shoda, Walter Mischel and Ozlem Ayduk, 'Preschoolers' delay of gratification predicts their body mass 30 years later' (2013) 162(1) *Journal of Pediatrics* 90.
[39] *Ibid*., 90; citing Angela L. Duckworth, Eli Tsukayama and Andrew B. Greier, 'Self-controlled children stay leaner in the transition to adolescence' (2010) 54(2) *Appetite* 304; Lori A. Francis and Elizabeth J. Susman, 'Self-regulation and rapid weight gain in children from age 3 to 12 years' (2009) 163(4) *Archives of Pediatrics and Adolescent Medicine* 297; Desiree M. Seeyave, Sharon Coleman, Danielle Appugliese, Robert F. Corwyn, Robert H. Bradley, Natalie S. Davidson, Niko Kaciroti and Julie C. Lumeng, 'Ability to delay gratification at age 4 years and risk of overweight at age 11 years' (2009) 163(4) *Archives of Pediatrics and Adolescent Medicine* 303.
[40] Schlam *et. al.* (2013), 91.
[41] B. J. Casey, Leah H. Somerville, Ian H. Gotlib, Ozlem Ayduk, Nicholas T. Franklin, Mary K. Askren, John Jonides, Marc G. Berman, Nicole L. Wilson, Theresa Teslovich, Gary Glover, Vivian Zayaz, Walter Mischel and Yuichi Shoda, 'Behavioral and neural correlates of delay of gratification 40 years later' (2011) 108(36) *Proceedings of the National Academy of Sciences* 14998.

performance on the original test as children and the performance on a go/no-go test as adults, a test also engaging the capacity for self-regulation. Second, Casey *et. al.* used fMRI to investigate whether structural differences could be found between subjects with higher and lower performance on the aforementioned tests. Subjects were classified as high or low delayers according to their childhood scores and completed the go/no-go task with hot and cold stimuli. The task requires subjects to respond to target stimuli and inhibit responses to non-targets; "hot" stimuli consisted of faces with emotional expressions which have previously been shown to bias behaviour.[42] The authors first found a difference between the two high- and low-delay groups only in the presence of emotional "hot" stimuli on the go/no-go test, with individuals who performed worse on the marshmallow test finding it consequently more difficult to supress responses to emotional cues on the go/no-go test. They comment that these findings are 'consistent with previous work suggesting that the capacity to resist temptation varies by context; the more tempting the choice for the individual, the more predictive are the individual differences in people's ability to regulate their behavior.'[43]

Turning to the second fMRI study, Casey *et. al.* found that the right inferior frontal gyrus was indicated in relation to accurately withholding responses; in comparison to subjects in the high delay group, lower delayers had correspondingly lower recruitment of this region on "no-go" relative to "go" trials. Additionally, there was a significant difference in activation of the ventral striatum with elevated activity for lower delayers relative to higher delayers; this region showed a three-way interaction between delay group, trial type and emotion, which 'highlights the importance of qualities of the stimulus people have to resist, such as its salience or allure, in modulating cognitive control ability.'[44] Taken together, the research by Casey *et. al.* provides a neurobiological basis for differences in the ability to resist impulses and temptation of immediate rewards in favour of longer-term goals. Moreover, the joint findings 'provide evidence that the ability to delay gratification assessed early in life predicts how well individuals can regulate

[42] *Ibid*., 14999; citing Todd A. Hare, Nim Tottenham, Matthew C. Davidson, Gary H. Glover and B. J. Casey, 'Contributions of amygdala and striatal activity in emotion regulation' (2005) 57(6) *Biological Psychiatry* 624; Leah H. Somerville, Todd Hare and B. J. Casey, 'Frontostriatal maturation predicts cognitive control failure to appetitive cues in adolescents' (2011) 23(9) *Journal of Cognitive Neuroscience* 2123.
[43] Casey *et. al.* (2011), 14999; citing Shoda, Mischel and Peake (1990).
[44] *Ibid*., 15000.

behavior years later, particularly when they are required to supress thoughts and actions toward alluring social cues.'[45]

A wealth of research has flowed from the original marshmallow test paradigm from which two general points might be made. First, the capacity for self-regulation has been demonstrated to be influential across a great number of faculties and situations, and impacts upon many areas of an individual's life, from educational attainment to body mass. As might be expected, a number of situational factors related to an individual's background and upbringing are demonstratively influential on the development of this capacity, from culture and socio-economic grouping to family upbringing and even trust in the researcher. Second, however, there is compelling evidence that the capacities for self-regulation developed in children as young as four have an ongoing impact throughout maturation and well into adulthood. These results must not be overstated and many of the correlations found have been modest; indeed, a huge range of individual factors will impact upon a person's educational attainment or relationship with food and weight throughout their lives. However, nor must the findings be understated; the fact that a general capacity measured at age four can have such statistically significant predictive power as many as four decades later strongly indicates that a person's capacities for self-regulation are notably determined during the earliest stages of life, from which point those capacities proceed to have an undeniably substantial influence over a disparate range of aspects of a person's life.

### *6.1.2. Critiques and Elaborations*

In addition to the original work and follow-up studies by Mischel *et. al.*, other research groups have similarly conducted both the original marshmallow test and other measures of self-control alongside subsequent longitudinal follow-up studies, shedding further light upon the predictive link between delay of gratification at preschool and performance across a disparate range metrics in adolescence and adulthood.[46] A further follow-up of

[45] *Ibid.*, 15001 – 15002.

[46] For example, see June P. Tangney, Roy F. Baumeister and Angie Luzio Boone, 'High self-control predicts good adjustment, less pathology, better grades, and interpersonal success' (2004) 72(2) *Journal of Personality* 271.

the original subjects was undertaken by Eigsti *et. al.* (with the participation of Mischel) to investigate whether their performance on the marshmallow test during childhood could similarly predict their performance on a go/no-go task.[47] Importantly, the authors did not find the predicted correlation between delay time as a child and performance on the go/no-go task. They suggest one possible explanation being that the marshmallow test relies not only on the subjects effectively controlling their attention (away from the tempting treat), but 'also on a number of other factors, such as motivation to obtain the delayed rewards.'[48] Indeed, in a similar 2011 study by Casey *et. al.*, above, correlation was found between the results of the marshmallow test and go/no-go test only on "hot" trials of the latter which used stimuli demonstrated to produce a motivational bias in adults. Alternatively, the relationship between attentional control and reaction time between the marshmallow test and the go/no-go tasks respectively may reflect a 'more general speed-of-processing ability rather than cognitive control.'[49]

Of greater significance, Watts, Duncan and Quan conducted a conceptual replication of Shoda, Mischel and Peake's 1990 work,[50] taking a sample of more than 900 children that was more representative of traits found across the general population such as race, ethnicity and parental education attainment. Such factors were then controlled for during analysis, observing their predictive quality on outcome variables of children recorded from age 54 months to 15 years and comparing these results with those from Shoda, Mischel and Peake's analysis. Later academic achievement and behaviour were modelled against delayed gratification at 54 months; models were subsequently tested that included controls for such characteristics as the home environment, earlier cognitive skills, and behavioural skills also assessed at 54 months old. The authors also found correlation, with each additional minute of delay in the marshmallow test at age 4 years predicting an increase in achievements at 15 years by approximately one-tenth of a standard deviation.

---

[47] Inge-Marie Eigsti, Vivian Zayas, Walter Mischel, Yuichi Shoda, Ozlem Ayduk, Mamta B. Dadlani, Matthew C. Davidson, J. Lawrence Aber and B. J. Casey, 'Predicting cognitive control from preschool to late adolescence and young adulthood' (2006) 17(6) *Psychological Science* 478.
[48] *Ibid*., 483.
[49] *Ibid*; citing Robert Kail and Timothy A. Salthouse, 'Processing speed as a mental capacity' (1994) 86(2/3) *Acta Psychologica* 199.
[50] Tyler W. Watts, Greg J. Duncan and Haonan Quan, 'Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes' (2018) 29(7) *Psychological Science* 1159.

Crucially, this correlation was between one half and two-thirds smaller than the size of that reported by Shoda, Mischel and Peake when controlling for such factors as family background, early cognitive abilities, and the home environment. Thus, factors such as socioeconomic background and other cognitive abilities displayed from a young age appear to be equally predictive of future achievement.

The conceptual replication by Watts, Duncan and Quan has not itself passed without notable critique. Falk, Kosse and Pinger conducted their own reanalysis of Watts' *et. al.* data with results yielding predictions closer to those reported by Mischel *et. al.*[51] Falk *et. al.* first contend that important measurement differences preclude a direct comparison of the results between Watts *et. al.* and Mischel *et. al.* For example, whereas subjects in the original marshmallow experiment had to wait as much as 15 minutes in order to receive the superior reward, the cut-off time for subjects in Watts' *et. al.* study was only 7 minutes. It is perfectly possible, however, that significant predictive data may have been gleaned from those subjects that were able to delay gratification for longer than 7 minutes in the original studies by Mischel *et. al*. Second, Falk *et. al.* submit that by controlling for covariates such as family background and early cognitive abilities, they may in fact be controlling for factors that contribute significantly towards the early ability to delay gratification. In this manner, the conceptual replication by Watts *et. al.* does not so much diminish the findings from Mischel *et. al.*, but provides 'suggestive evidence that early environment shapes a child's ability to delay gratification… in accordance with a body of related evidence.'[52]

This latter point is explored further by Doebel, Michaelson and Munakata who argue that 'many of the variables in [Watts' *et. al.*] models should not have been included as confounds because they likely captured factors that measure fundamental processes

[51] Armin Falk, Fabian Kosse and Pia Pinger, 'Re-revisiting the marshmallow test: A direct comparison of studies by Shoda, Mischel and Peake (1990) and Watts, Duncan, and Quan (2018)' (2019) *Psychological Science* 1.

[52] Falk, Kosse and Pinger (2019), 4; citing Thomas Deckers, Armin Falk, Fabian Kosse, Pia Pinger and Hannah Schildberg-Hörisch, 'Socio-economic status and inequalities in children's IQ and economic preferences' (IZA Institute of Labor Economics, Discussion paper no. 11158, November 2017); Armin Falk and Fabian Kosse, 'Early childhood environment, breastfeeding and the formation of preferences' (SOEP Papers on multidisciplinary panel data research 882-2016).

supporting delay of gratification.'[53] Specifically, Watts *et. al.* included two sets of covariates across two sets of models – child background, home environment, general cognitive skills and behavioural skills – their justification being that the first two are unlikely to be the targets of early childhood interventions whilst the latter two are unlikely to be targeted by interventions focussing on the 'narrow set of skills involved with gratification delay.'[54] However, Doebel *et. al.* contest that these variables in fact measure processes underlying the delay of gratification and, indeed, are reasonable targets for interventions. For example, Watts *et. al.* controlled for executive functions which have otherwise been theorised to support delayed gratification through maintaining longer-term goals and inhibiting impulses.[55] They also controlled for early verbal abilities which have similarly been theorised to support executive functions and demonstrated through moderate correlations.[56]

Further, Doebel *et. al.* contest that a number of the factors related to a child's background and family environment that were controlled as covariates by Watts *et. al.* play a similarly important role in supporting delayed gratification, both in the moment and across developmental time. These include 'social norms, values, and trust, which may influence children's tendency to exercise delay of gratification both developmentally and when they are confronted with temptation.'[57] For example, theoretical and empirical work supports the view that parenting and language may 'scaffold self-regulatory skills that children use

[53] Sabine Doebel, Laura E. Michaelson and Yuko Munakata, 'Good things come to those who wait: Delaying gratification likely does matter for later achievement (A commentary on Watts, Duncan, & Quan, 2018)' (2019) *Psychological Science* 1, 1.
[54] *Ibid*.
[55] *Ibid*; citing Adele Diamond, 'Executive functions' (2013) 64(1) *Annual Review of Psychology* 135; Akira Miyake and Naomi P. Friedman, 'The nature and organization of individual differences in executive functions: Four general conclusions' (2012) 21(1) *Current Directions in Psychological Science* 8.
[56] Laura J. Kuhn, Michael T. Willoughby, Lynne Vernon-Feagans, Clancy B. Blair and Family Life Project Key Investigators, 'The contributions of children's time-specific and longitudinal expressive language skills on developmental trajectories of executive function' (2016) 148 *Journal of Experimental Child Psychology* 20; Stephanie M. Carlson and Louis J. Moses, 'Individual differences in inhibitory control and children's theory of mind' (2001) 72(4) *Child Development* 1032.
[57] Doebel, Michaelson and Munakata (2019), 2; citing Stephanie M. Carlson and Philip David Zelazo, 'The value of control and the influence of values' (2011) 108(41) *Proceedings of the National Academy of Sciences* 16861; Bettina Lamm, Heidi Keller, Johanna Teiser, Helene Gudi, Relindis D. Yovsi, Claudia Freitag, Sonja Poloczek, Ina Fassbender, Janina Suhrka, Manuel Teubert, Isabel Vöhringer, Monika Knoopf, Gudrum Schwarzer and Arnold Lohaus, 'Waiting for the second treat: Developing culture-specific models of self-regulation' (2018) 89(3) *Child Development* e261.

when they need to delay gratification'[58] whilst such processes have also become the target of some early childhood interventions.[59] Thus, Doebel *et. al.* argue, the analysis by Watts *et. al.* has controlled for a whole range of factors which may themselves fully support the ability to delay gratification demonstrated in the marshmallow test. Having excluded such factors directly relevant to delaying gratification, 'the weakened link between early delay of gratification and later outcomes is not surprising.'[60]

The crucial point that may be taken from Watts' *et. al.* partial conceptual replication of Mischel's seminal work is that, where the self-control *per se* that a person is able to exercise at age four is indeed reliably predictive of a range of metrics in the future, this capacity for self-regulation is itself comprised of, or supported by, a whole range of other factors, from that individual's socio-economic background to their family environment, and other cognitive functions similarly expressed at a young age. Each of these factors have a role to play in self-control, for which performance on the marshmallow test is a potent indicator. Considering the divergent findings between Mischel *et. al.* and Watt *et. al*, Michaelson and Munakata[61] conducted their own secondary analysis of the data and found significant correlations for three of the five outcomes tested, including relationships between the ability to delay gratification in childhood and problem behaviour later on. They write:

> 'These relationships were better explained by social support than by self-control, suggesting that the marshmallow test is predictive because it

[58] *Ibid*; citing Annie Bernier, Stephanie M. Carlson, Marie Deschênes and Célie Matte-Gagné, 'Social factors in the development of early executive functioning: A closer look at the caregiving environment' (2012) 15(1) *Developmental Science* 12; Stuart I. Hammond, Ulrich Müller, Jeremy I. M. Carpendale, Maximilian B. Bibok and Dana P. Liebermann-Finestone, 'The effects of parental scaffolding on preschoolers' executive function' (2012) 48(1) *Developmental Psychology* 271; Lynne Vernon-Feagans, Michael Willoughby, Patricia Garrett-Peters and Family Life Project Key Investigators, 'Predictors of behavioral regulation in kindergarten: Household chaos, parenting, and early executive functions' (2016) 52(3) *Developmental Psychology* 430.

[59] Adele Diamond, W. Steven Barnett, Jessica Thomas and Sarah Munro, 'Preschool program improves cognitive control' (2007) 318(5855) *Science* 1387.

[60] Doebel, Michaelson and Munakata (2019), 1.

[61] Laura E. Michaelson and Yuko Munakata, 'Same data set, different conditions: Preschool delay of gratification predicts later behavioral outcomes in a preregistered study' (2020) *Psychological Science* 1.

> reflects aspects of a child's early environment that are important over the long term.'[62]

And, indeed, there is theoretical and empirical research abound to illustrate the crucial, if not inescapable influence that such aspects of a child's early environment and personality will exert throughout their overall development, subsequent decision-making, achievements and failures. Highlighting some examples of evidence in this direction, a significant study by Lamm *et. al.*[63] replicating the marshmallow paradigm across German and Cameroonian children revealed cultural differences impacting upon children's ability to delay gratification, making similar findings to the original work by Mischel.[64] Kidd, Palmeri and Aslin demonstrate how performance on the marshmallow task can be moderated by the child's beliefs concerning the reliability of their environment whilst,[65] in a similar vein, Ma, Chen, Xu, Lee and Heyman show that a child's level of trust in the researcher can equally moderate how long they delay gratification.[66] Finally, controlling for variables concerning early achievement, demography and home environment, Ahmed, Tang, Waters and Davis-Kean show that working memory at four-and-a-half years is significantly predictive of achievement at age 15.[67] These examples further demonstrate how the capacity for self-regulation is comprised of, or supported by, other components of early cognitive abilities such as working memory.

### *6.1.3. The Legal Relevance of the Marshmallow Test and Self-Regulation*

In many ways, Mischel and his colleagues foresaw the revelations of Watts' *et. al.* partial conceptual replication of the marshmallow test, alongside the body of research that has flown therefrom demonstrating the influence and predictive power of different factors on

[62] *Ibid*., 1.
[63] Lamm *et. al.* (2018).
[64] Mischel (1958).
[65] Celeste Kidd, Holly Palmeri and Richard N. Aslin, 'Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability' (2013) 126(1) *Cognition* 109.
[66] Fengling Ma, Biyun Chen, Fen Xu, Kang Lee and Gail D. Heyman, 'Generalized trust predicts young children's willingness to delay gratification' (2018) 169 *Journal of Experimental Child Psychology* 118.
[67] Sammy F. Ahmed, Sandra Tang, Nicholas E. Waters and Pamela Davis-Kean, 'Executive function and academic achievement: Longitudinal relations from early childhood to adolescence' (2019) 111(3) *Journal of Educational Psychology* 446.

self-regulation and long-term achievement. Mischel *et. al.* considered that the association between delayed gratification and adolescent competencies might at least partially reflect the operation of "cognitive construction competencies." From this viewpoint:

> '[T]he *qualities that underlie* effective self-imposed delay in preschool *may be crucial ingredients* of an expanded construct of "intelligent social behavior" that encompasses social as well as intellectual knowledge, coping, and problem-solving competencies.'[68]

Thus, it was envisaged from the outset that the capacity for self-regulation demonstrated through the marshmallow test may itself be comprised of, or facilitated by, a range of underlying factors, both endogenous such as a child's performance on other cognitive tests at that same age, and external factors such as that child's upbringing, household, family and socio-economic background.

The point that is particularly relevant to the discussion of legal responsibility is how demonstrably engrained a capacity for self-regulation appears to be from such a young age, and determined by factors entirely outside of any child's individual control. Caution is necessary once again – it is not stated that a person's expression of self-control in any particular circumstance at age 30 years has been entirely and irrevocably determined by their performance on a marshmallow test at age four years. Plainly, the self-control that anybody exerts in a particular moment or situation is determined by the entirety of factors leading up to that moment, and not merely their cognitive abilities as a child. Nevertheless, the totality of the evidence surrounding the marshmallow test indicates that a person's *general* capacity for self-regulation is heavily determined from an early age by the *whole range of factors* which support that general capacity, from other cognitive abilities such as working memory, to exogenous influences such as culture, socio-economic status, family and upbringing.

Again, this does not mean that any individual is necessarily predestined at age four to commit criminal acts when they are an adult. However, owing to the prevalent link

[68] Shoda, Mischel and Peake (1990), 985.

between criminality and deficiencies in self-regulation, this does mean that an individual's *predisposition* to commit a criminal offence in adulthood *because of* a deficiency in self-control, is undoubtedly and inescapably shaped by factors from that individual's childhood and early development. How strongly deficiencies in the development of self-regulation during childhood determine a person's self-control in adulthood remains an open question. Nonetheless, those differences between people who can exert greater versus lesser self-control by age four years are strong enough to predict visible structural differences in the brain networks related to self-control for those same subjects more than four decades later. Finally, it is trite to observe that the entire range endogenous and exogenous factors that may contribute towards or facilitate self-control in a child must be entirely outside of that child's influence or control. It is parents, families, schools and society more generally who together bear the responsibilty of controlling these factors towards the healthy and flourishing development of all children.

With this in mind, it is submitted that it is incumbent upon any model of (legal) responsibility to take into consideration the causality of a defendant's criminal behaviour and conduct. This will be reflected in two aspects of legal responsibility in particular. On the one hand, the law recognises that certain causes of loss of control may provide a partial or total defence, such as through the eponymous partial defence of "loss of control" and the complete defence of automatism respectively. In this manner the law recognises that, in certain circumstances, a defendant's loss of control may excuse their subsequent criminal behaviour. On the other hand, the causes of a person's criminality may become a yet more relevant consideration when a convicted defendant comes to be sentenced. The law currently does take such factors into account, for example, when making orders for some form of medical treatment, therapy or rehabilitation. However, it is submitted that such rehabilitative theories of punishment ought to be paramount in circumstances where an individual's criminal conduct has resulted from even a diminished capacity for self-control caused by identifiable, demonstrable, and compelling factors outside of an individual's sphere of influence. These ideas are explored more fully in chapter twelve of this thesis, below.

## 6.2. ***The Conscious Veto***

In his seminal work, discussed above in section 5.2 of this thesis, Benjamin Libet revealed preparatory brain activity in the form of the "readiness potential" or "RP" which was produced before conscious awareness of a decision to act and yet appeared to predict the outcome of that decision. This paradigm has been replicated consistently and, whilst the RP is no longer considered to represent that actual timing of a decision, further experiments discussed in chapters three and five have revealed brain activity predictive of decision outcomes recorded several seconds before any reported conscious awareness of a decision to act. In Libet's original findings, however, a small window of some 150 milliseconds remained after the moment when subjects reported becoming aware of their decisions to act and within the available remaining time for that decision to be inhibited. Libet referred to this potential as the conscious veto,[69] which has colloquially become known as "free won't" in contrast to the "free will" which Libet's work has widely been interpreted as precluding.

Such "free won't" – the decision to inhibit an otherwise intended and prepared action – is the component of a decision that is arguably of greatest significance to the question of legal responsibility. It is the choice of *whether* or not to commit a particular criminal action that ultimately converts a simple *mens rea* into a coincidence of *mens rea* with *actus reus* and, consequently, criminal liability. Discussed in section 2.1.3 of this thesis, above, Brass and Haggard present a modified version of the Libet paradigm within an fMRI study,[70] providing evidence for the involvement of the dorso-fronto-median cortex and the anterior insula in decisions to veto or inhibit a prepared voluntary action. The question then follows, if decisions to act are reached by the brain before arising to the level of conscious awareness, as suggested by the body of evidence flowing from the

[69] Benjamin Libet, 'Do we have free will?' (1999) 6(8-9) *Journal of Consciousness Studies* 47, 51 – 52.
[70] Marcel Brass and Patrick Haggard, 'To do or not to do: The neural signature of self-control' (2007) 27(34) *Journal of Neuroscience* 9141; Simone Kühn, Patrick Haggard and Marcel Brass, 'Intentional inhibition: How the "veto-area" exerts control' (2009) 30(9) *Human Brain Mapping* 2834; see also Adam R. Aron, Trevor W. Robbins and Russell A. Poldrack, 'Inhibition and the right inferior frontal cortex: One decade on' (2014) 18(4) *Trends in Cognitive Sciences* 177; Giovanni Mirabella, 'Endogenous inhibition and the neural basis of "free won't"' (2007) 27(51) *Journal of Neuroscience* 13919.

Libet paradigm, is the same true for the separate component of the decision of *whether* or not to veto a particular intention?

Evidence suggests that a great many different perceptual and motor processes can and do operate outside of consciousness, such as provided by subliminal priming[71] and patient studies.[72] Consequently, many authors like Libet have posited that cognitive control functions require consciousness – *i.e.,* those associated with the prefrontal cortices, including conflict detection and response inhibition.[73] Van Gaal, Ridderinkhof, Wildenberg and Lamme explain:

> '[T]he logic behind the consciousness-control relationship is the idea that we usually become aware of stimuli that interfere or interrupt routine action, which are the same stimuli that call for adaptive control operations. Therefore, it has been proposed that higher level control operations, such as response inhibition, *depend* on the conscious detection of response-relevant warning signals. Following this line of reasoning, it should not be possible to trigger inhibitory control processes when the instruction stimulus itself is presented subliminally.'[74]

[71] Stanislas Dehaene, Lionel Naccache, Gurvan le Clec'H, Etienne Koechlin, Michael Mueller, Ghislaine Dehaene-Lambertz, Pierre-François van de Moortele and Denis le Bihan, 'Imaging unconscious semantic priming' (1998) 395(6702) *Nature* 597; Martin Eimer and Friederike Schlaghecken, 'Effects of masked stimuli on motor activation: Behavioral and electrophysical evidence' (1998) 24(6) *Journal of Experimental Psychology* 1737; Dirk Vorberg, Uwe Matler, Armin Heinecke, Thomas Schmidt and Jens Schwarzbach, 'Different time courses for visual perception and action priming' (2003) 100(10) *Proceedings of the National Academy of Sciences* 6275.

[72] For example, see Jon Driver and Jason B. Mattingley, 'Parietal neglect and visual awareness' (1998) 1(1) *Nature Neuroscience* 17; Lawrence Weiskrantz, 'Blindsight revisited' (1996) 6(2) *Current Opinion in Neurobiology* 215.

[73] For example, see Libet (1999); Stanislas Dehaene and Lionel Naccache, 'Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework' (2001) 79(1/2) *Cognition* 1; Bernard J. Baars, 'The conscious access hypothesis: Origins and recent evidence' (2002) 6(1) *Trends in Cognitive Sciences* 47; Martin Eimer and Friederike Schlaghecken, 'Response facilitation and inhibition in subliminal priming' (2003) 64(1/2) *Biological Psychology* 7.

[74] Simon van Gaal, Richard Ridderinkhof, Wery P. M. van den Wildenberg and Victor A. F. Lamme, 'Dissociating consciousness from inhibitory control: Evidence for unconsciously triggered response inhibition in the stop-signal task' (2009) 35(4) *Journal of Experimental Psychology* 1129, 1129; citing Dehaene and Naccache (2001); Eimer and Schlaghecken (2003).

The authors set out to investigate this question through a modified version of the "stop-signal" paradigm;[75] in this test, subjects must perform a rapid right- or left-hand button press on go signals whilst, on a small proportion of trials, the go signal is proceeded by a stop signal following which subjects must refrain from responding. Subjects are more likely to respond to the stop signal when it is presented shortly after the go signal, whereas the greater the time between the two signals, the closer the go process reaches towards completion and the less likely subjects are to inhibit their response. Manipulating the "stop-signal delay" ('SSD') therefore provides an estimate of the stop-signal reaction time ('SSRT') – *i.e.,* the duration of the inhibitory process. Van Gaal *et. al.* varied the paradigm by masking the stop signals optimally and sub-optimally, rendering them invisible and visible respectively. It was hypothesised that if it is possible to trigger response inhibition unconsciously, this ought to be displayed through small differences in inhibition rates for the masked stop signals.[76]

The authors found that task-relevant signals attended and processed unconsciously can indeed 'actively trigger and initiate response inhibition, thereby breaking the alleged intimate relationship between consciousness and inhibitory control.'[77] The results further demonstrated that cognitive control functions could be differentially affected by conscious awareness; thus, whereas online inhibitory control operations (*i.e.,* immediate inhibition) could be triggered unconsciously, strategic trial-by-trial control operation (*e.g.,* slowing responses after committing errors) was not so triggered in masked stop-signal conditions.[78] Moreover, inhibitory control was demonstrably less efficient when unconsciously triggered and processed; 'although non-masked stop signals lead to complete response inhibition on the majority of trials, this is the exception rather than the rule on trials containing a masked stop signal.'[79] Whilst inhibition could therefore clearly be triggered and processed entirely unconsciously, this unconscious response was generally less flexible, less efficient, and slower than when consciousness was engaged.

---

[75] Gordon D. Logan, 'On the ability to inhibit thought and action: A users' guide to the stop signal paradigm' in Carr D. D. T. H. (ed.), *Inhibitory Processes in Attention, Memory and Language* (Academic Press 1994).
[76] Gaal *et. al.* (2009), 1129 – 1130; see also Simon van Gaal, Richard Ridderinkhof, H. Steven Scholte and Victor A. F. Lamme, 'Unconscious activation of the prefrontal no-go network' (2010) 30(11) *Journal of Neuroscience* 4143.
[77] *Ibid*., 1135.
[78] *Ibid*.
[79] *Ibid*., 1135 – 1136.

This finding can be interpreted following the general proposition throughout this thesis that consciousness operates to enhance and improve existing underlying functions – such as decision-making or response inhibition – by providing greater time and mental resources to those operations. Whereas self-regulation can be engaged and processed entirely unconsciously, consciousness can and does improve that capacity, lending it more time and greater resources with which to operate.

Having determined that decisions to veto an otherwise intended action *can* arise and operate unconsciously, the further question follows how the timing of such (unconscious) decisions to veto an action relates to the timing of subjective conscious awareness of that veto decision. Kühn and Brass[80] began to explore this question with a modification of the go/no-go paradigm which introduced an additional condition; after a go signal, one-quarter of trials were followed by either a stop-signal or a decide signal, the latter indicating that the subject should themselves freely choose whether or not to proceed with an otherwise prepared action. The rationale follows that it should be possible to compare the reaction times between trials where subjects voluntarily decided to press the button with those in which they were instructed to do so. Equally, individuals ought to be able to subjectively distinguish between when they have done something impulsively (when there was no time to deliberate) and when they have consciously decided to veto that action.[81]

Trials in which the subjects responded to a go-signal impulsively, or in which a decide-signal was too late and the subjects had already enacted a quick impulse response, appeared with faster reaction times of ~600 milliseconds. Trials in which the subjects successfully responded to a stop-signal or decided to inhibit actions in response to a decide-signal showed no response time; and trials in which subjects successfully responded to a decide-signal but proceeded to press the button appeared with a comparatively slower reaction time of ~1400 milliseconds (***figure i***).

[80] Simone Kühn and Marcel Brass, 'Retrospective construction of the judgment of free choice' (2009) 18(1) *Consciousness and Cognition* 12.
[81] *Ibid*., 13.

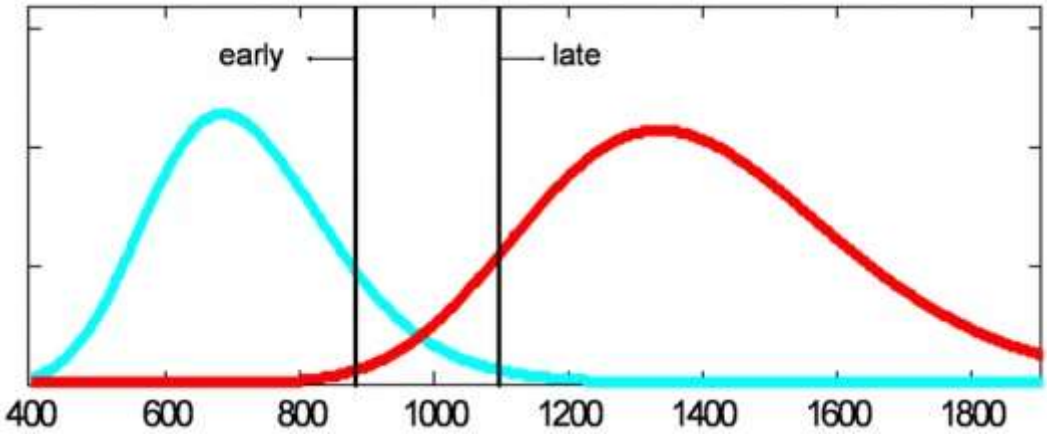

***Fig. i – schematic drawing of criterion to assign trials to early- and late-decision categories.***[82]

However, the finding of particular note was that when the subjects were later questioned regarding their decisions, they were unanimously 'convinced that they stopped the prepared action and went through a decision process in the early decide-go trials but actually did not.'[83] That is to say, in trials where the reaction times indicated that the decision *must* have been taken quickly and without opportunity to inhibit and then recontinue a prepared action, subjects nonetheless reported choosing to press the button as a result of a deliberative process – they were apparently unable to detect the fact that they had not actually paused a prepared action and then decided to re-initiate it, which would have been impossible within the timeframe of the early-go decisions. Kühn and Brass write:

> 'This clearly argues against Libet's assumption that a veto process can be consciously initiated. He used the veto in order to reintroduce the possibility to control the unconsciously initiated actions. But since the subjects are not very accurate in observing when they have not stopped, the act of vetoing cannot be consciously initiated.'[84]

One year later, Walsh, Kühn, Brass, Wenke and Haggard further investigated the *whether* component of decision-making in a study broadly replicating the original Libet

[82] *Ibid.*, 16.
[83] *Ibid.*, 17.
[84] *Ibid.*, 20.

paradigm.[85] In particular, subjects were to prepare a voluntary action to make a key press whilst reporting the timing of their subjective awareness of reaching such a decision. In some trials, however, subjects were to freely decide to inhibit the button press at the last moment, with EEG recordings being compared between the different action and inhibition trials. The action trials displayed stereotypical patterns of a reduced beta-band spectral power prior to the movement followed by a rebound after the movement.[86] The inhabitation trials displayed significantly different activity, however, consisting of an increase in spectral power indicated at the left frontal hemisphere and peaking at 12 milliseconds prior to the subjectively reported intention to move.[87] That is to say, a clearly distinct EEG signature was recorded for inhibiting prepared actions, the peak of which occurred *prior to subjective awareness of forming the intention to move which would later be inhibited*.

Given that the aforementioned finding falls to a distinction of 12 milliseconds, there is more than ample room for error. In 2013, however, Filevich, Kühn and Haggard revisited question of the timing of the conscious veto in a further EEG study.[88] Subjects were required to either make a rapid button press or temporarily inhibit that response; in this way, the researchers 'operationalized inhibition as a transient process, characterised by delayed responding, rather than as a complete suppression of all behavioural output.'[89] On each trial subjects could be given one of five instructions: to make a rapid button press, to make a delayed button press, to make a free choice whether to act rapidly, to make a free choice whether to act after a delay, or not to act at all. The rationale follows that neural networks 'continually exhibit small fluctuations in state, which may have significant effects on behaviour… [and] may be particularly relevant for behaviour in the absence of other clear, strong external signals.'[90] The aim, therefore, was to identify the

[85] Eamonn Walsh, Simone Kühn, Marcel Brass, Dorit Wenke and Patrick Haggard, 'EEG activations during intentional inhibition of voluntary action: An electrophysiological correlate of self-control?' (2010) 48(2) *Neuropsychologia* 619.
[86] *Ibid*., 622.
[87] *Ibid*.
[88] Elisa Filevich, Simone Kühn and Patrick Haggard, 'There is no free won't: Antecedent brain activity predicts decisions to inhibit' (2013) 8(2) *PLoS ONE* e53053; see also Elisa Filevich, Simone Kühn and Patrick Haggard, 'Intentional inhibition in human action: The power of "no"' (2012) 36(4) *Neuroscience & Biobehavioral Reviews* 1107.
[89] *Ibid*., e53053.
[90] *Ibid*., e53054.

potential effects of such fluctuations in trials where subjects made a "free" choice between immediate or delayed action.

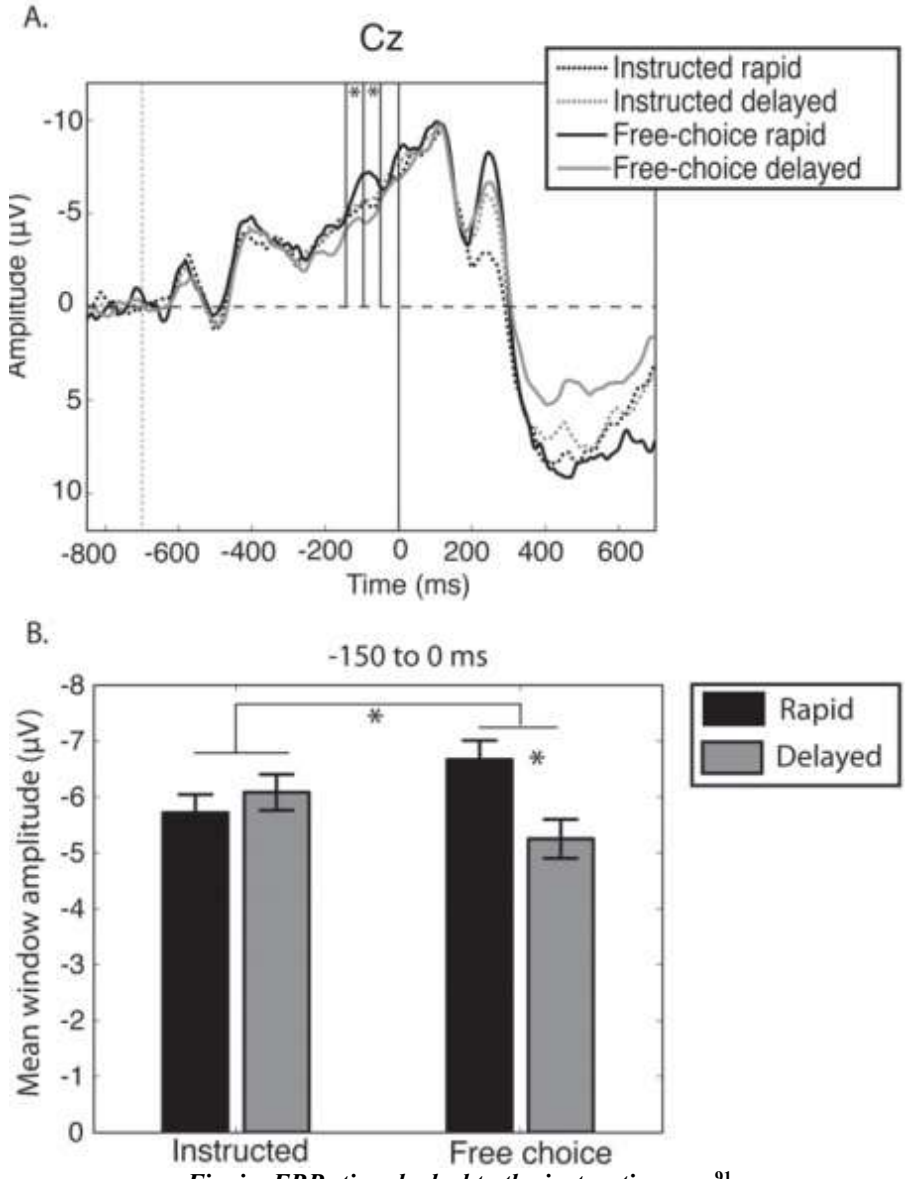


***Fig. j – ERPs time-locked to the instruction cue.***[91]

Crucially, EEG recordings were taken in time intervals of –150 to –100 milliseconds, –100 to –50 milliseconds, and –50 to 0 milliseconds *prior* to subjects receiving their instruction as to which type of trial they were in. Consequently, the subjects *cannot* have known whether or not they would be compelled to act or have a free choice between acting immediately or after a delay at the times when the critical EEG recordings were

[91] *Ibid.*, e52058.

being taken. Filevich, Kühn and Haggard found significant differences in activity on electrode Cz at the time windows of –150 to –100 milliseconds, and –100 to –50 milliseconds prior to instruction. Specifically:

> 'In the free choice condition, the [contingent negative variation ('CNV')] amplitude measured from Cz was reduced (*i.e.*, less negative) when participants chose to transiently inhibit and delay than when they chose to respond rapidly. In contrast, the instructed condition showed no difference between rapid and delay trials. That is, the CNV amplitude just before the decision cue had a specific association with subsequent free choices to respond rapidly or to delay.'[92]

The authors conclude their main argument submitting that the results provide, for the first time, a candidate for Libet's "conscious veto" which, as with conscious decisions to act in the first place, results from preceding unconscious neural activity of which consciousness later becomes aware.

In further follow-up work, Parkinson and Haggard conducted a modified go/no-go task where subjects were required to make a rapid button press in response to a go signal, withhold responses on a no-go signal, or make a free and spontaneous choice between both options on a decision-signal.[93] However, prior to the instruction on each trial, subjects were subliminally primed with masked primes which could be either congruent or incongruent with the go or no-go signals. A measure of response times and the proportion of action choices in the free decision-signal provided the key variables, whilst primes were presented at latencies which would either be positively or negatively compatible, following previous literature.[94] They found that masked go-signal primes provided at positively compatible latencies did increase the speed of response times as

---

[92] *Ibid*.

[93] Jim Parkinson and Patrick Haggard, 'Subliminal priming of intentional inhibition' (2014) 130(2) *Cognition* 255.

[94] *Ibid*., 256; citing Odmar Neumann and Werner Klotz, 'Motor responses to nonreportable, masked stimuli: Where is the limit of direct parameter specification?' in Umiltà C. and Moscovitch M. (eds.), *Attention and Performance 15: Conscious and Nonconscious Information Processing* (Massachusetts Institute of Technology Press 1994); Martin Eimer, 'Facilitatory and inhibitory effects of masked prime stimuli on motor activation and behavioural performance' (1999) 101(2/3) *Acta Psychologica* 293.

expected, but did not influence the number of instances that subjects chose to act on free trials. However, masked go-signal primes provided at negatively compatible latencies significantly increased decisions to inhibit on free trials.

The 2014 results from Parkinson and Haggard suggest that, contrary to previous supposition,[95] 'relatively late, volitional, high-level cognitive control processes can be manipulated by non-conscious means.'[96] A further 2015 follow-up work by the same authors replicated the modified version of the go/no-go task whilst introducing EEG analysis.[97] In trials where subjects were instructed to inhibit a prepared button press, stereotypical EEG activity was recorded at the N2 component that has previously been associated with specific no-go responses, whilst similar activity was not recorded when subjects were instructed with the go-signal.[98] Of greatest interest, however, is that the N2 component displayed the stereotypical activity for inhibition on the free decision trials for *both* an intentional decision to act and to inhibit action. Parkinson and Haggard propose that an ordinary volitional decision to act begins with the first step of inhibiting prepotent responses to a particular choice before generating an actual decision; in this regard, 'intentional inhibition has a crucial role breaking the flow of stimulus-driven responding, allowing expression of volitional decisions.'[99]

Concluding generally on the role of consciousness in both "free will" and "free won't", Haggard writes forcefully:

> '[T]here is *no convincing evidence* that this intentional inhibition is a "*conscious* veto", in the sense of a brain-independent conscious cause. Just

[95] For example, Stanislas Dehaene, Eric Artiges, Lionel Naccache, Catherine Martelli, Armelle Viard, Franck Schürhoff, Christophe Recasens, Marie Laurie Paillère Martinot, Marion Leboyer and Jean-Luc Martinot, 'Conscious and subliminal conflicts in normal subjects and patients with schizophrenia: The role of the anterior cingulate' (2003) 100(23) *Proceedings of the National Academy of Sciences* 13722.
[96] Parkinson and Haggard (2014), 263.
[97] Jim Parkinson and Patrick Haggard, 'Choosing to stop: Responses evoked by externally triggered and internally generated inhibition identify a neural mechanism of will' (2015) 27(10) *Journal of Cognitive Neuroscience* 1948.
[98] *Ibid*., 1951 – 1953; citing Hirokazu Bokura, Shuhei Yamaguchi and Shotai Kobayashi, 'Electrophysiological correlates for response inhibition in a Go/NoGo task' (2001) 112(12) *Clinical Neurophysiology* 2224; Michael Falkenstein, Jöorg Hoormann and Joachim Hohnsbein, 'ERP components in Go/NoGo tasks and their relation to inhibition' (1999) 101(2/3) *Acta Psychologica* 267.
[99] Parkinson and Haggard (2015), 1948.

> as the experience of conscious will is in fact a consequence of preceding brain activity, our sense of "conscious veto" must also be a consequence of unconscious brain activity. Intentional inhibition therefore involves a specific set of brain processes, which both prevent the prepared action, and produce the conscious experience of inhibition. The conscious experience itself, however, does not cause anything.'[100]

Whether or not this statement underplays the overall involvement of consciousness in deciding first what to do in a given situation and second whether or not to carry out that intention into action, the evidence surrounding both the original Libet paradigm and subsequent replications and, indeed, the experiments considered throughout this thesis, overwhelmingly indicates towards a significantly diminished role for consciousness than that which is typically assumed. With regards to the *whether* component of a decision specifically, the experiments explored in this chapter similarly indicate that the final and critical decision of whether or not to enact a particular intention is reached first by the brain outside of conscious awareness, which arises second and, therefore, cannot be the source of self-regulation in any particular moment. As previously stated, this does not preclude a role for consciousness in improving decision-making capacities – including the *whether* component of a decision and the capacity for self-control – by directing greater mental resources to those capacities and allowing more time for them to operate.

### *6.2.1. The Legal Relevance of the (Un)conscious Veto*

The research discussed in the present section arguably has the greatest implications for theories of punishment discussed further in chapter twelve, below. The previous sections 3.2 and 5.2.4 of this thesis presented a range of experiments flowing from the original Libet paradigm which provide compelling empirical support for the proposition that the brain reaches a decision of *what* to do prior to individuals becoming consciously aware of that decision outcome. The same is reasoned to hold even when a decision is consciously deliberated over time; each option considered and the evaluation given

[100] Patrick Haggard, 'Neuroethics of free will' in Illes J. and Sahakian B. J. (eds.), *The Oxford Handbook of Neuroethics* (Oxford University Press 2011), 222.

thereto is the product of prior unconscious cerebral activity that subsequently arises to the level of conscious awareness. However, Libet's original work left open the door for the possibility that consciousness may have a critical role to play in vetoing decided actions and controlling impulses, and for much time hence it was reasonably assumed that consciousness was a prerequisite for exercising executive functions of self-regulation.

The studies considered in this chapter are submitted to provide similarly compelling evidence that consciousness is, after all, *not* a necessary prerequisite of self-control. In particular, just as neural correlates have been reliably identified that can accurately anticipate the content and timing of decisions to do some particular action, so corresponding correlates have been identified for the act of inhibiting or vetoing a prepared decision. Most importantly, the neural signatures are similarly found to arise prior to conscious awareness even of the decision to act that is to be subsequently vetoed. In any sequence of conscious deliberation – the weighing of options, the positive selection of a particular option, and the subsequent inhibition of enacting that selection – it is submitted that each substantive decision within that sequence is first reached unconsciously before an individual becomes consciously aware of the content of that decision. Applying the proposition that consciousness improves the functioning of existing mental processes by lending greater time and mental resources to those operations, it is perfectly reasonable to appreciate why conscious deliberation can dramatically improve the outcome of any decision-making process, despite the fact that each stage in that process is determined first unconsciously, and enters into conscious awareness second.

The above notwithstanding, it remains the case that just as the decisions of *what* to do, *how* to do it, and *when*, can be triggered, processed and initiated into action by the brain entirely automatically and outside of conscious awareness, so too can the final and critical decision of *whether* or not to implement a plan into physical action. Considering the subjective mental states which form the *mens rea* of a criminal offence, such as intention, dishonesty or recklessness, it has previously been argued in section 3.3 of this thesis that the existence of such mental states does not provide *proof* that that mental state has been arrived at consciously and deliberately by an agent, as opposed to arising automatically

as the result of purely unconscious processing. The evidence considered in the present chapter extends the same argument to the *whether* component of a decision – *i.e.,* the fact that an individual has proceeded to commit a certain (criminal) act is not *alone* proof that they have consciously and deliberately chosen to proceed with that act, as it is quite possible that their brain unconsciously and automatically decided not to veto the relevant bodily actions, without input or awareness from consciousness altogether. Equally, where somebody embarks upon some criminal conduct but then changes their mind, this *alone* is not proof that they have consciously and deliberately chosen not to proceed with that act, as it is quite possible that their brain unconsciously and automatically decided to veto the relevant bodily actions. Indeed, the evidence in this chapter suggests that the initiation of such a veto is indeed triggered unconsciously by the brain, with conscious awareness of that veto decision arising later.

Philosophically, the implications of this argument extend even deeper. Suppose, as argued in section 3.3 of this thesis, that a decision to do *x* is arrived at first unconsciously, with conscious awareness thereof arising second. Next, the unconscious brain decides to veto *x* in favour of *y*, with conscious awareness thereof, again, arising second. After further deliberation, the unconscious brain decides to veto *y* and returns in favour of *x*, and conscious awareness follows. Finally, suppose that the individual is on the verge of implementing *x* into action, but the brain unconsciously generates a final veto signal and the individual holds themselves back. The individual's subjective experience may be one of deliberating between two options, *x* and *y*, and of perhaps changing their mind therebetween. However, the evidence considered in chapters three to six of this thesis suggest that each stage of the deliberation – deciding to *x*, vetoing *x*, deciding to *y*, vetoing *y*, *etc.* – are first decided unconsciously, with conscious awareness thereof arising second. It may well be the case that the periods of conscious awareness throughout the deliberation process do in fact make a causal contribution to the decision outcome – the present thesis has proposed that consciousness may improve the deliberative process as such by providing time and greater mental resources to that process.

Yet, it remains the case that it is not consciousness *per se*, nor direct, online, conscious *control* which "forces" the brain to switch from decision *x* to *y* and back again, nor which

forces the brain to veto the final decision *x*. As the evidence in section 5.2 of this thesis explored, conscious processes are generally slow, and conscious awareness of stimuli – whether exogenous phenomena in the world or endogenous thoughts, sensations and feelings – only arises after a certain threshold and duration of neural excitement has been established. Meanwhile, the brain networks involved in deciding *what* to do, *how* to do it, *when* and *whether* to do it and, ultimately, *why* to do it, operate continuously, automatically and in parallel, whether or not a particular decision is being contemplated consciously. It is those same processes which continue to operate and provide the solutions to each component of a decision when that decision is being consciously deliberated; however, the fact of conscious deliberation is itself unlikely to radically alter the way those networks operate nor the automatic processes involved. Rather, as above, it is postulated that the contribution of consciousness lies in giving those processes more time over which to operate and accumulate evidence and, further, devotes greater mental resources to those processes when they are the focus of conscious attention.

With the above in mind, it is further submitted that retributive theories of punishment cannot reasonably be justified. Retributivism rests fundamentally upon the principle that people deserve to be punished because, in the moment of committing any given criminal act, they had the capacity and opportunity to choose to do otherwise. This is premised upon the notion that people are in direct conscious control of their decisions and actions, contrary to the suggestion of the present and preceding chapters of this thesis. Rather, in the particular moment of enacting a criminal intention, both the decision to act and the absence of a decision to veto that action are determined by brain activity outside of the *direct conscious* control of the individual. Even when criminal plans have been consciously deliberated over time, the fact that an individual does not exhibit the brain activity require to veto that action in the moment of commission lies outside of that person's direct conscious control. Consequently, the proposition that a person can consciously choose to do otherwise in the particular moment of committing a criminal offence is arguably undermined, and with it the retributive theory of punishment.

## 6.3. ***The Strength Model of Self-Control***

The research considered so far would tend to give the impression that self-regulation is some immutable faculty over which people have no influence or control. And, indeed, the evidence considered in the immediately preceding section of this thesis does strongly suggest that the self-control that a person exerts (or, as the case may be in the event, does not exert) in any particular moment is triggered into operation outside of conscious awareness and, therefore, conscious control. That notwithstanding, there is a growing body of evidence indicating that a person's *capacity* for self-control is changeable after all and, crucially, strategies may be adopted (even in adulthood) that appear to improve this general self-regulatory capacity. It is important to note that such improvements in self-control do not equate to granting people more "direct" or conscious influence over self-control in any particular moment of action; it remains the case that inhibitory responses most likely are triggered outside of consciousness. Nevertheless, a general capacity for self-regulation may be exercised and improved such that, in any given situation, an individual's unconscious processes of self-regulation will function better.

Just as this thesis is founded upon the premise that all human behaviour and action is determined, not least including conduct that is subject to criminal sanction, so future behaviour may be determined, to a lesser or greater degree, by positive interventions in a person's life. Rehabilitative and deterrent theories of punishment rest upon this very assertion – that a person's future conduct may be purposefully directed away from criminality by certain interventions. Whilst it is readily accepted that the criminal justice system is far from the ideal forum within which such interventions are made to improve upon anybody's self-regulatory capacities, the clear link between failures in self-control and criminality render the legal system a necessary forum within which such interventions must inevitably take place. What is more, with the capacity for self-control being one of the crucial mental capacities underlying the legal concept of volition, it is entirely appropriate that the law responds to criminal failures in self-control by attempting to utilise any reasonable and viable means to improve an offender's capacity for self-control, thereby reforming and rehabilitating them.

Arguably one of the leading models of how self-control functions in the brain is the "Strength Model", which draws an analogy between mental self-control and a physical muscle.[101] In a similar manner, both can be exerted and require a certain degree of energy to do so; what is more, both muscles and self-control can become tired from exertion, and 'acts of self-control cause short-term impairments ("ego depletion") in subsequent self-control, even on unrelated tasks.'[102] Equally like a muscle, emerging evidence suggests that the capacity for self-control can also be improved through regular exercise, with such improvements again being demonstrated across unrelated tasks and domains whilst, similarly, 'blood glucose is an important component of the energy' required to exert and maintain self-control.[103]

Thus, just as exercise can strengthen leg muscles resulting in improved performance in both running and jumping, so exercising self-control can deliver benefits across a range of domains, for example, from improving dieting and resistance to unhealthy foods to improving financial habits and resistance to impulsive spending, *etc*. As with any physical training, there almost certainly exist engrained limits as to how far any individual's capacity for self-regulation might be influenced and improved in adulthood, such limits having been determined by the range of factors discussed throughout this chapter. Whilst arguably anybody's capacity for self-control might be improved, therefore, there remain 'stable individual differences in self-regulation'[104] that no amount of training and exercise would be able to entirely modify.

[101] See Roy F. Baumeister, Matthew T. Gailliot, C. Nathan DeWall and Megan Oaten, 'Self-regulation and personality: How interventions increase regulatory success, and how depletion moderates the effects of traits on behavior' (2006) 74(6) *Journal of Personality* 1773; Roy F. Baumeister, Kathleen D. Vohs and Dianne M. Tice, 'The strength model of self-control' (2007) 16(6) *Current Directions in Psychological Science* 351.

[102] Baumeister, Vohs and Tice (2007), 351; see further Mark Muraven, Dianne M. Tice and Roy F. Baumeister, 'Self-control as limited resource: Regulatory depletion patterns' (1998) 74(3) *Journal of Personality and Social Psychology* 774; Roy F. Baumeister, Ellen Bratslavsky, Mark Muraven and Dianne M. Tice, 'Ego depletion: Is the active self a limited resource?' (1998) 74(5) *Journal of Personality and Social Psychology* 1252.

[103] *Ibid*; on the other hand, the ability to train and improve self-regulation is contentious – for example, see Brian M. Lee and Markus Kemmelmeier, 'How reliable are the effects of self-control training?: A re-examination using self-report and physical measures' (2017) 12(6) *PLoS ONE* e0178814.

[104] Baumeister, Gailliot, DeWall and Oaten (2006), 1174.

The concept of ego depletion has been demonstrated across a variety of circumstances, lending support to the significant influence that capacities for self-regulation have across a range of domains.[105] For example, in studies where subjects have depleted some of their self-regulatory "energy" on a first task, their performance is poorer in subsequent tasks, capturing the effects of ego depletion on reasoning about difficult problems;[106] relying on more simplistic decision-making strategies; [107] being more prone to impulsive spending;[108] demonstrating higher levels of aggression;[109] consuming more alcohol even when there is an impending good reason not to do so;[110] exerting less control over, and being more likely to perform, inappropriate sexual behaviours;[111] being more prone to break diets;[112] and generally 'present[ing] themselves in ways less likely to make a good impression.' [113] In a similar vein, research is beginning to elucidate the types of behaviours which consume the mental energy required for maintaining self-control and thus lead to ego depletion. Examples include the effort of making decisions,[114] interacting

[105] *Ibid*., 1176; see also Mark Muraven and Roy F. Baumeister, 'Self-regulation and depletion of limited resources: Does self-control resemble a muscle?' (2000) 126(2) *Psychological Bulletin* 247.
[106] Brandon J. Schmeichel, Kathleen D. Vohs and Roy F. Baumeister, 'Intellectual performance and ego depletion: Role of the self in logical reasoning and other information processing' (2003) 85(1) *Journal of Personality and Social Psychology* 33.
[107] Anastasiya Pocheptsova, On Amir, Ravi Dhar and Roy F. Baumeister, 'Deciding without resources: Resource depletion and choice in context' (2009) 46(3) *Journal of Marketing Research* 344.
[108] Ronald J. Faber and Kathleen D. Vohs, 'To buy or not to buy? Self-control and self-regulatory failure in purchase behavior' in Baumeister R. F. and Vohs K. D. (eds.), *Handbook of Self-Regulation: Research, Theory, and Applications* (The Guildford Press 2004).
[109] Tanja S. Stucke and Roy F. Baumeister, 'Ego depletion and aggressive behavior: Is the inhibition of aggression a limited resource?' (2006) 36(1) *European Journal of Social Psychology* 1.
[110] Mark Muraven, R. Lorraine Collins and Kristen Nienhaus, 'Self-control and alcohol restraint: An initial application of the self-control strength model' (2002) 16(2) *Psychology of Addictive Behaviors* 113.
[111] Matthew T. Gailliot and Roy F. Baumeister, 'Self-regulation and sexual restraint: Dispositionally and temporarily poor self-regulatory abilities contribute to failures at restraining sexual behavior' (2007) 33(2) *Personality and Social Psychology Bulletin* 173.
[112] Kathleen D. Vohs and Todd F. Heatherton, 'Self-regulatory failure: A resource-depletion approach' (2000) 11(3) *Psychological Science* 249.
[113] Baumeister, Gailliot, DeWall and Oaten (2006), 1176; citing Kathleen D. Vohs, Roy F. Baumeister and Nathalie Ciarocco, 'Self-regulation and self-presentation: Regulatory resource depletion impairs impression management and effortful self-presentation depletes regulatory resources' (2005) 88(4) *Journal of Personality and Social Psychology* 632.
[114] Kathleen D. Vohs, Roy F. Baumeister, Jean M. Twenge, Brandon J. Schmeichel, Dianne M. Tice and Jennifer Crocker, 'Decision fatigue exhausts self-regulatory resources – But so does accommodating to unchosen alternative' (2005) (unpublished) <https://www.researchgate.net/publication/237738528_Decision_Fatigue_Exhausts_Self-Regulatory_Resources_-_But_So_Does_Accommodating_to_Unchosen_Alternatives> accessed 29 January 2021.

with others regarding whom a person holds negative and derogatory opinions,[115] and effortfully presenting oneself in a manner that is unusual and non-habitual.[116]

The first indications that the general capacity for self-regulation could be improved through exercise and training arose from an experiment concerning ego depletion by Muraven, Baumeister and Tice.[117] First, a measure of ego depletion was taken using a standard procedure; subjects perform a hand-grip stamina task to provide a baseline, followed by a thought suppression task where they must refrain from thinking about a white bear,[118] followed again by a second hand-grip stamina task. Subjects typically displayed reduced stamina after the suppression task as a result of ego depletion. Second, subjects were assigned one of three exercises to be performed over two weeks – tracking what food they ate, improving their mood, or improving their posture – whilst a control group had no exercises to perform. Third, the subjects returned to the laboratory after two weeks to repeat the first stamina procedure measuring ego depletion. In concurrence with the strength model of self-regulation, subjects that had performed regular self-regulation exercises performed significantly better on the hand-grip task than controls (albeit only for the tasks of tracking food consumption and improving their posture). What is more, those subjects who followed the training exercises most consistently displayed the best improvements in self-control.

Yet stronger evidence for the training and improvement of self-regulatory capacities was provided in a set of experiments by Oaten and Cheng.[119] In the first study, subjects signed up to a two-month physical exercise course which included a range of resistance and cardiovascular training in a program specifically designed by a specialist for each individual subject. Self-regulation was measured before and after the two-month training

[115] Jennifer A. Richeson and J. Nicole Shelton, 'When prejudice does not pay: Effects of interracial contact on executive function' (2003) 14(3) *Psychological Science* 287.
[116] Vohs, Baumeister and Ciarocco (2005).
[117] Mark Muraven, Roy F. Baumeister and Dianne M. Tice, 'Longitudinal improvement of self-regulation through practice: Building self-control strength through repeated exercise' (1999) 139(4) *Journal of Social Psychology* 446.
[118] Following a paradigm by Daniel M. Wegner, David J. Schneider, Samuel R. Carter and Teri L. White, 'Paradoxical effects of thought suppression' (1987) 53(1) *Journal of Personality and Social Psychology* 5.
[119] Megan Oaten and Ken Cheng, 'Longitudinal gains in self-regulation from regular physical exercise' (2006b) 11(4) *British Journal of Health Psychology* 717; Megan Oaten and Ken Cheng, 'Improvements in self-control from financial monitoring' (2007) 28(4) *Journal of Economic Psychology* 487.

period through a number of measures, including a visual tracking task where subjects must resist shifting their attention to a nearby comedic distractor.[120] After the training program, the effects of ego depletion had clearly, strongly and substantially diminished. What is more, the improvements in self-control were manifested in a range of other domains, including reductions in cigarette, alcohol, caffeine, junk food consumption and impulsive spending, and improvements in emotional control and attention to activities such as studying rather than watching television. As Baumeister, Gailliot, DeWall and Oaten comment, these changes 'suggest an across-the-board improvement in self-control, consistent with the strength model.'[121]

The second study by Oaten and Cheng followed a similar concept;[122] subjects enrolled in a four-month financial monitoring program which included a personalised spending review, money management plan, spending diary and other logs. Whilst the subjects' incomes did not increase over the four months, their average improvement was to increase their rate of saving from around 8% to 38% of income. More relevant, the subjects also improved on the various measures of self-regulation, including the visual tracking task described above, again, in addition to a 'variety of positive side effects indicative of a central improvement in self-regulatory strength.' [123] These included a reduced consumption of cigarettes, alcohol, caffeine and unhealthy food, and improvements in emotional control, the maintenance of chores and commitments, and study habits. A third similar study by Oaten and Chen garnered similar results where a study intervention program improved self-regulatory strength and dampened the effects of stress induced by academic exams, as a well as eliciting a similar range of other benefits.[124]

[120] Following a modified paradigm by Zenon W. Pylshyn and Ron W. Storm, 'Tracking multiple independent targets: Evidence for a parallel tracking mechanism' (1988) 3(3) *Spatial Vision* 179.
[121] Baumeister, Gailliot, DeWall and Oaten (2006), 1782.
[122] Oaten and Cheng (2007).
[123] Baumeister, Gailliot, DeWall and Oaten (2006), 1783.
[124] Megan Oaten and Ken Cheng, 'Improved self-control: The benefits of a regular program of academic study' (2006a) 28(1) *Basic and Applied Social Psychology* 1.

A number of further studies have utilised different methods of both measuring and improving self-regulatory strength;[125] 'once again, the self-control exercises made people subsequently less vulnerable to ego depletion.'[126] In addition, further evidence in favour of the strength model of self-control has been provided from a meta-analysis of existing studies.[127] As Baumeister, Gailliot, DeWall and Oaten conclude, the discussed studies suggest that it is possible to improve self-regulation through exercise following the strength model. Moreover, the experiments 'suggest that improving self-regulation operates by increasing a general core capacity. That is, as the person performs exercises to improve self-regulation in one sphere, he or she becomes better at self-regulating in other spheres.'[128]

### *6.3.1. Self-Control Depletion and Criminal Behaviour*

The links between learning disabilities, poor mental health and criminal behaviour are by now well documented. The overrepresentation of people with mental health disorders can be found in countries all around the world;[129] there is a significantly higher prevalence of mental health disorders, self-harm and suicide within UK prisons than in the general population,[130] with some 37% of NHS spending on adult healthcare in prisons being devoted to mental health and substance abuse issues.[131] Moreover, estimates vary from some 20% to 50% or more of prisoners having some form of learning or developmental disability,[132] many of which patently diminish a person's ordinary capacities for self-

[125] Matthew T. Gailliot, E. Ashby Plant, David A. Butz and Roy F. Baumeister, 'Increasing self-regulatory strength can reduce the depleting effect of suppressing stereotypes' (2007) 33(2) *Personality and Social Psychology Bulletin* 281.
[126] Baumeister, Gailliot, DeWall and Oaten (2006), 1786.
[127] Martin S. Hagger, Chantelle Wood, Chris Stiff and Nikos L. D. Chatzisarantis, 'Ego depletion and the strength model of self-control: A meta-analysis' (2010) 136(4) *Psychological Bulletin* 495.
[128] Baumeister, Gailliot, DeWall and Oaten (2006), 1783.
[129] For example, see Seena Fazel, Adrian J. Hayes, Katrina Bartellas, Massimo Clerici and Robert Trestman, 'The mental health of prisoners: A review of prevalence, adverse outcomes and interventions' (2016) 3(9) *Lancet Psychiatry* 871; Seth J. Prins, 'The prevalence of mental illnesses in U.S. State prisons: A systematic review' (2014) 65(7) *Psychiatric Services* 862.
[130] Miriam Light, Eli Grant and Kathryn Hopkins, *Gender Differences in Substance Misuse and Mental Health Amongst Prisoners: Results from the Surveying Prisoner Crime Reduction (SPCR) Longitudinal Cohort Study of Prisoners* (Ministry of Justice 2013), 17 – 20.
[131] House of Commons Committee of Public Accounts, *Mental Health in Prisons* (HC 400, Eighth Report of Session 2017-19), 8.
[132] See Jenny Talbot and Chris Riley, 'No one knows: Offenders with learning difficulties and learning disabilities' (2007) 35(3) *British Journal of Learning Disabilities* 154, 156.

control, such as attention deficit hyperactivity disorder ('ADHD'), bipolar disorder and schizophrenia, to cite some obvious examples. From such data it may be inferred that a 'large number of inmates have deficiencies in self-control.'[133] Furthermore, conditions of poverty are similarly overrepresented within prison populations, for example, with some 15% of UK convicts being homeless immediately prior to incarceration.[134] This is particularly significant in light of the vicious cycle between poverty and poor self-control highlighted in the introduction to the present chapter – the third step in that cycle, it appears, is criminality.

Indeed, perhaps one of the strongest assertions of an intractable link between self-regulation and crime is provided in Gottfredson and Hirschi's seminal *A General Theory of Crime*.[135] To summarise their argument, it is observed that most criminal offences are relatively simple to execute, do not require any particular long-term planning and offer few long-term benefits in return. With this in mind, the authors further highlight that individuals with deficiencies or deficits in self-control would be more prone to impulsivity, risk-taking, short-sightedness and insensitivity to others and, consequently, would be more likely to be involved in criminal behaviour. However, the authors go much further than this, arguing that deficiencies in self-control are *the* causes of not only crime but a range of analogous detrimental behaviours including drug and alcohol abuse, gambling and poorer social relationships (the "generality" postulate). Finally, Gottfredson and Hirschi submit that the capacity for self-control is substantively developed during childhood (the "origins" postulate) and thereafter remains relatively stable throughout the rest of an individual's life (the "stability" postulate).

Many of the assertions made by Gottfredson and Hirschi have since been demonstrated through empirical evidence, not least including the range of studies explored in the present chapter which have each spoken to one or more of the three aforementioned

[133] Polaris Koi, Susanne Uusitalo and Jarno Tuominen, 'Self-control in responsibility enhancement and criminal rehabilitation' (2018) 12(2) *Criminal Law and Philosophy* 227, 236.
[134] *Ibid*., citing Kim Williams, Jennifer Poyser and Kathryn Hopkins, *Accommodation, Homelessness and Reoffending of Prisoners: Results from the Surveying Prisoner Crime Reduction (SPCR) Survey* (Ministry of Justice 2012).
[135] Michael R. Gottfredson and Travis Hirschi, *A General Theory of Crime* (Stanford University Press 1990); see further Michael R. Gottfredson and Travis Hirschi, *Modern Control Theory and the Limits of Criminal Justice* (Oxford University Press 2020).

postulates. Furthermore, gathering the large body of experimental evidence surrounding the topic of self-regulation, two notable meta-analyses have specifically set out to test the empirical status of Gottfredson and Hirschi's "General Theory." Pratt and Cullen published one such meta-analysis in 2000, reviewing 21 empirical studies integrating almost 50,000 individual cases.[136] First, they found a consistently significant effect of low self-control as a predictor of criminal behaviour which, when compared with concurrent literature at the time, 'would rank self-control as one of the strongest known correlates of crime.'[137] Second, the effects of self-control appeared to be general, impacting not only upon crime but analogous behaviours also. However, the meta-analysis did not reveal any particular support for the stability postulate; for example, social learning variables also exerted a strong effect, suggesting that these may impact upon self-control beyond childhood and adolescence.[138] Nonetheless, the analysis provides a compelling indication that Gottfredson and Hirschi's 'core proposition that low self-control increases involvement in criminal and analogous behaviors is empirically supported.'[139]

A further meta-analysis by Engel in 2012 included more than 100 empirical studies and similarly found the theoretical link between poor self-control and criminality to be 'overwhelmingly supported.'[140] In particular, he found that self-control significantly explained the frequency and intensity of deviance in 88% of the empirical studies reviewed and, with the exception of 4 out of 717 cases, that effect was always significant and inverse; lower self-control predicted greater deviance.[141] However, the analysis further demonstrated how the effects of self-control on deviance may also be significantly moderated by certain factors, including age, culture and employment status.[142] It is noteworthy that these are many of the same factors that have previously been identified as being closely linked to, and influential over, the development of the capacity for self-control in the first place during childhood. Thus, whilst self-control has an undeniable

[136] Travis C. Pratt and Francis T. Cullen, 'The empirical status of Gottfredson and Hirschi's General Theory of Crime: A meta-analysis' (2000) 38(3) *Criminology* 931.
[137] *Ibid*., 951 – 952.
[138] *Ibid*., 952 – 953.
[139] *Ibid*., 953.
[140] Christoph Engel, 'Low self-control as a source of crime: A meta-study' (Reprints of the Max Planck Institute for Research on Collective Goods, Bonn 2012/4), 24.
[141] *Ibid*.
[142] *Ibid*., 25.

link to criminal behaviour, it naturally follows that so, too, do many of those factors indicated as significantly influencing ordinary capacities for self-control.[143]

*

One vital implication flowing from the research discussed concerns the functioning of normal capacities of self-regulation in a typical, "reasonable" and law-abiding person. The previous section highlighted the clear overrepresentation of people within prison populations who might reasonably be inferred as suffering from deficiencies in self-control by reason of a demonstrative and identifiable pathology. That is to say, the link between both mental illnesses and learning / developmental deficiencies, their impact upon ordinary self-control, and resultant criminal behaviour is well-established. But identifying certain classes of people – *i.e.,* the mentally ill and / or disabled – as having appreciably diminished capacities for self-control presupposes that healthy, neurotypical adults possess some unwaveringly higher capacity. In reality, the strength model of self-regulation implies that the ordinary capacities of healthy, reasonable people fluctuate as a result of ego depletion. Whilst, therefore, it may be entirely correct to say that one or both of the average and maximal capacities for self-control possessed by people with certain mental illnesses or disabilities is markedly lower than that of neurotypical people, the self-control exerted by the latter group may nevertheless diminish to levels more readily observed in the former group as a result of ego depletion. Thus, ego depletion may have a significant impact on the self-control and, therefore, likelihood of criminality for even the ordinary, healthy, reasonable person on the street.

Gino, Schweitzer, Mead and Ariely provide an illuminating set of studies exploring how ego depletion may promote unethical behaviour by rendering people unable to resist temptations.[144] In the first study, testing the hypothesis that ego depletion increases the

[143] See further Alexander T. Vazsonyi, Jakub Mikuška and Erin L. Kelley, 'It's time: A meta-analysis on the self-control – deviance link' (2017) 48 *Journal of Criminal Justice* 48; Alex R. Piquero, John MacDonald, Adam Dobrin, Leah E. Daigle and Francis T. Cullen, 'Self-control, violent offending, and homicide victimization: Assessing the general theory of crime' (2005) 21(1) *Journal of Quantitative Criminology* 55.
[144] Francesca Gino, Maurice E. Schweitzer, Nicole L. Mead and Dan Ariely, 'Unable to resist temptation: How self-control depletion promotes unethical behavior' (2011) 115(2) *Organizational Behavior and Human Decision Processes* 191.

prevalence of unethical behaviour,[145] subjects first watched a video clip consisting of a woman being interviewed whilst irrelevant words were displayed under the screen.[146] Test subjects were instructed to ignore the words thereby expending self-regulatory resources, whilst controls were provided with no such instruction. This was followed by a cheating assessment task in which subjects undertook a problem-solving task with the opportunity to falsely report a higher score in order to earn more money.[147] As predicted, subjects in the ego depletion condition rated the video as significantly more difficult to follow than controls and, in the event, more than twice as many subjects from the ego depletion condition overstated their performance on the cheating test.[148]

The second study tested the hypotheses that ego depletion would result in reduced moral awareness,[149] and that the 'impaired ability to recognize moral issues mediates the relationship between self-regulatory resource depletion and unethical behavior.'[150] Subjects were first required to complete an essay-writing task with those in the ego depletion condition being required to avoid using words containing the letters "A" and "N".[151] Subjects then performed a similar cheating test where they had the opportunity to overreport their own scores in order to earn more money. Finally, subjects completed a word-completion task where they had to convert word fragments into meaningful words, comparing how many ethics-related to non-ethics-related words were produced by each

[145] *Ibid*., 192; citing Mark Muraven, Greg Pogarsky and Dikla Shmueli, 'Self-control depletion and the general theory of crime' (2006) 22(3) *Journal of Quantitative Criminology* 263; Nicole L. Mead, Roy F. Baumeister, Francesca Gino, Maurice E. Schweitzer and Dan Ariely, 'Too tired to tell the truth: Self-control resource depletion and dishonesty' (2009) 45(3) *Journal of Experimental Social Psychology* 594.

[146] Following Daniel T. Gilbert, Douglas S. Krull and Brett W. Pelham, 'Of thoughts unspoken: Social interference and the self-regulation of behavior' (1988) 55(5) *Journal of Personality and Social Psychology* 685.

[147] Following Nina Mazar, On Amir and Dan Ariely, 'The dishonesty of honest people: A theory of self-concept maintenance' (2008) 45(6) *Journal of Marketing Research* 633.

[148] Gino, Schweitzer, Mead and Ariely (2011), 194.

[149] *Ibid*., 193; citing Mark D. Street, Scott C. Douglas, Scott W. Geiger and Mark J. Martinko, 'The impact of cognitive expenditure on the ethical decision-making process: The cognitive elaboration model' (2001) 86(2) *Organizational Behavior and Human Decision Processes* 256.

[150] *Ibid*; citing Karl Aquino and Americus Reed, 'The self-importance of moral identity' (2002) 83(6) *Journal of Personality and Social Psychology* 1423.

[151] Following Brandon J. Schmeichel, 'Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control' (2007) 136(2) *Journal of Experimental Psychology* 241.

group.[152] Once again, more than twice as many subjects in the depletion condition cheated as those in the control group. Additionally, ego depleted subjects produced fewer words related to ethics and morality on the word-completion task than controls did, 'suggesting that self-regulatory resource depletion reduced moral awareness.'[153] What is more, the analysis showed that subjects' moral awareness 'mediated the relationship between self-regulatory resource depletion and cheating.'[154]

The third study by Gino, Schweitzer, Mead and Ariely tested the hypothesis that moral identity would further mediate the relationship between ego depletion and unethical behaviour, such that individuals with a stronger moral identity would be less susceptible to the effects of depletion on their ethical behaviour.[155] Subjects completed a replication of the resource-depletion and cheating tasks with the addition of a questionnaire including measures of moral identity.[156] Previous results were again replicated, whilst the additional questionnaire confirmed that 'moral identity weakens the association between depletion and unethical behavior.'[157] Finally, a fourth study which similarly provided subjects with an opportunity to cheat before measuring self-control revealed that the very act of resisting the temptation to cheat depletes self-regulatory resources. The studies are important for demonstrating how even a relatively minor depletion of resources such as through completing a trivial distraction task can have a significant impact upon a person's likelihood of proceeding to engage in dishonest and unethical behaviour, namely cheating to make a higher financial gain. Moreover, the research lends further support to the appreciation of immorality and, by extension, criminality as being products of depleted self-control as proposed by Gottfredson and Hirschi's general theory of crime.

---

[152] Following Francesca Gino and Max H. Bazerman, 'When misconduct goes unnoticed: The acceptability of gradual erosion in others' unethical behavior' (2009) 45(4) *Journal of Experimental Social Psychology* 708.
[153] Gino, Schweitzer, Mead and Ariely (2011), 197.
[154] *Ibid*.
[155] See also Yan Wang, Guosen Wang, Qiuju Chen and Lin Li, 'Depletion, moral identity, and unethical behavior: Why people behave unethically after self-control exertion' (2017) 56 *Consciousness and Cognition* 188.
[156] Following Aquino and Reed (2002).
[157] Gino, Schweitzer, Mead and Ariely (2011), 198.

A number of studies have continued to further illuminate the effects of ego depletion on ordinary people in ways that may readily be appreciated to increase the likelihood of falling into potentially criminal behaviour. Friehe and Schildberg-Hörisch investigate the effects of self-regulatory depletion on risk-taking and antisocial behaviour,[158] both of which have clear relations to crimes of recklessness and violence respectively. Subjects were depleted using an established task where they must cross out particular letters in a text according to certain rules regarding their position in a word.[159] Risk-taking was measured using similarly established paradigms in which subjects decide how many points to invest in an exchange with a one-half probability of yielding a dividend of 2.5 times the investment.[160] Investments below the maximum number of points provide a measure of risk aversion, and can be exchanged for cash at the end of the experiment. Equally, anti-social behaviour was measured using an established game where subjects have the option of acting more or less antisocially towards an opponent.[161] The study did find a significant effect of self-control on risk-taking, but did not identify a similar effect on antisocial behaviour.[162] More specifically, 'while low trait self-control is positively correlated with antisocial behavior, a reduction in the current level of self-control causes a slight, however not significant decrease in antisocial behavior.'[163]

Whist Friehe and Schildberg-Hörisch did not find a specific significant effect of self-control depletion on antisocial behaviour, DeWall, Finkel and Denson review a number of studies which do strongly suggest that depletion of self-regulatory resources can be a significant predictor of aggression towards both romantic partners and strangers.[164] For

[158] Tim Friehe and Hannah Schildberg-Hörisch, 'Self-control and crime revisited: Disentangling the effect of self-control on risk taking and antisocial behavior' (DICE Discussion paper no. 264, 2017).
[159] Following Baumeister, Bratslavsky, Muraven and Tice (1998).
[160] Following Uri Gneezy and Jan Potters, 'An experiment on risk taking and evaluation periods' (1997) 112(2) *Quarterly Journal of Economics* 631; Gary Charness and Uri Gneezy, 'Strong evidence for gender differences in risk taking' (2012) 83(1) *Journal of Economic Behavior & Organization* 50.
[161] Following Armin Falk and Urs Fischbacher, '"Crime" in the lab – Detecting social interaction' (2002) 46(4/5) *European Economic Review* 859.
[162] See also William P. McClanahan and Sander van der Linden, 'An uncalculated risk: Ego-depletion reduces the influence of perceived risk but not state affect on criminal choice' (2020) *Psychology, Crime & Law* 1.
[163] Friehe and Schildberg-Hörisch (2017), 20 – 22.
[164] C. Nathan DeWall, Eli J. Finkel and Thomas F. Denson, 'Self-control inhibits aggression' (2011) 5(7) *Social and Personality Psychology Compass* 458.

example, DeWall, Baumeister, Stillman and Gailliot[165] conducted one study where resource-depleted subjects received an insulting evaluation of an essay that they had written,[166] following which they had the opportunity to exact revenge by making the evaluator unwittingly eat food with hot sauce. As predicted, those subjects with depleted self-control displayed more aggressive behaviour by pouring more hot sauce on the food; moreover, these finding were replicated over further measures of aggression, such as punishing the confederate with a blast of loud noise or a negative job evaluation. Demonstrating effects of self-regulation in the opposite direction DeWall, Finkel and Denson further cite a study by Denson, Capper, Oaten, Friese and Schofield which showed that improving or enhancing self-control can correspondingly reduce aggression.[167] In a similar vein, the authors cite further evidence that self-control failures can predict violence against intimate partners.[168]

Finally, a series of studies concerning the role of self-regulatory failures for unethical conduct in the workplace has obvious potential relevance for many white-collar crimes and criminal offences committed by companies by way of their directors and employees. For example,[169] research by Welsh and Ordóñez suggests that consecutive performance goals such as are highly familiar within the workplace environment can deplete self-regulatory resources and consequently exacerbate unethical behaviour over time.[170] In a

[165] C. Nathan DeWall, Roy F. Baumeister, Tyler F. Stillman and Matthew T. Gailliot, 'Violence restrained: Effects of self-regulation and its depletion on aggression' (2007) 43(1) *Journal of Experimental Social Psychology* 62; see also Stucke and Baumeister (2006).

[166] Following Brad J. Bushman and Roy F. Baumeister, 'Threatened egoism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence?' (1998) 75(1) *Journal of Personality and Social Psychology* 219.

[167] DeWall, Finkel and Denson (2011), 462; citing Thomas F. Denson, Miriam M. Capper, Megan Oaten, Malte Friese and Timothy P. Schofield, 'Self-control training decreases aggression in response to provocation in aggressive individuals' (2011) 45(2) *Journal of Research in Personality* 252.

[168] DeWall, Finkel and Denson (2011), 463 – 464; citing Eli J. Finkel, C. Nathan DeWall, Erica B. Slotter, Megan Oaten and Vangie A. Foshee, 'Self-regulatory failure and intimate partner violence perpetration' (2009) 97(3) *Journal of Personality and Social Psychology* 483.

[169] For overview, see Russell E. Johnson, Szu-Han Lin and Hun W. Lee, 'Self-control as the fuel for effective self-regulation at work: Antecedents, consequences, and boundary conditions of employee self-control' (2018) 5 *Advances in Motivation Science* 87.

[170] David T. Welsh and Lisa D. Ordóñez, 'The dark side of consecutive high performance goals: Linking goal setting, depletion, and unethical behavior' (2014) 123(2) *Organizational Behavior and Human Decision Processes* 79; see also Lisa D. Ordóñez and David T. Welsh, 'Immoral goals: How goal setting may lead to unethical behavior' (2015) 6 *Current Opinion in Psychology* 93; Maurice E. Schweitzer, Lisa D. Ordóñez and Bambi Douma, 'Goal setting as a motivator of unethical behavior' (2004) 47(3) *Academy of Management Journal* 422.

similar vein, Mitchell *et. al.* present evidence that performance pressure encourages cheating within the workplace, identifying both anger and self-serving cognitions as mediators of performance pressure on cheating behaviour.[171] Indeed, similar findings have been identified amongst students enrolled in business school where a lack of self-control and a desire to become rich were associated with a higher propensity to engage in unethical conduct when faced with temptations.[172] In relation to anger and antisocial behaviours specifically, one study by Barnes, Lucianetti, Bhave and Christian reveals that poor quality (but not quantity) of sleep may significantly contribute to ego depletion resulting in the abusive treatment of employees by their depleted supervisor.[173]

### *6.3.2. The Legal Relevance of Self-Control Depletion*

None of the immediately aforementioned studies actually show subjects engaging in criminal behaviour as a direct result of ego depletion; setting up such an experiment would be patently unethical. What these experiments *do* demonstrate is the effect that a depletion of self-control can have on behaviours such as cheating, dishonesty and aggression towards others in healthy neurotypical individuals. Within the right context, however, cheating, dishonesty, aggression and abuse can all readily inflate into a criminal action – perhaps cheating on taxes, being dishonest and stealing at a shop checkout, or lashing out verbally or physically towards another.

As is discussed in considerably more detail in chapter eight of this thesis, below, the law presumes that all adult defendants are sane and possess the capacity for conscious self-control, *unless and until* sufficient evidence is raised to the contrary. Consequently, the fact that a defendant was possessed of self-control is not a matter that must ordinarily be proven by the prosecution, but may often be sought to be negated by the defence. This is

---

[171] Marie S. Mitchell, Michael D. Baer, Maureen L. Ambrose, Robert Folger and Noel F. Palmer, 'Cheating under pressure: A self-protection model of workplace cheating behavior' (2018) 103(1) *Journal of Applied Psychology* 54.

[172] Godfred Matthew Yaw Owusu, Rita Amoah Bekoe, Theodora Aba Abekah Koomson and Samuel Nana Yaw Simpson, 'Temptation and the propensity to engage in unethical behaviour' (2018) 35(1) *International Journal of Ethics and Systems* 43.

[173] Christopher M. Barnes, Lorenzo Lucianetti, Devasheesh P. Bhave and Michael S. Christian, '"You wouldn't like me when I'm sleepy": Leaders' sleep, daily abusive supervision, and work unit engagement' (2015) 58(5) *Academy of Management Journal* 1419.

particularly relevant to formal defences such as diminished responsibility, loss of control, (insane) automatism, duress, necessity and self-defence, each of which are concerned to some degree with whether circumstances exist that are recognised as impacting to a greater or lesser degree upon any ordinary person's capacity for self-control. Often, therefore, the crucial question will be whether the circumstances were such that the self-control of the hypothetical ordinary reasonable person would fairly be regarded as being sufficiently diminished or abrogated entirely so as to satisfy one of the aforementioned defences.

The vital point to extrapolate from the evidence considered in this section, therefore, is to appreciate that the capacity for self-regulation of even the hypothetical ordinary reasonable person is not a fixed and unchanging capacity but one which may be significantly depleted by any number of factors at any given point in time. The hypothetical reasonable person encompasses the range of capacities for self-regulation that may be encountered amongst the broad spectrum of "reasonable" people that make up society, from the most diligent puritan to the more impulsive immediate gratifier. It is only when conduct falls below a lower boundary of what might be expected from the hypothetical reasonable person that a person's actions may be regarded as criminal; nobody is expected to act as a paragon of virtue and self-control. Following the strength model of self-regulation and the theory of ego depletion, however, it *must* be recognised that that lower bound of reasonable self-control itself fluctuates and may be reduced as particular circumstances dictate.

The practical implication of this follows, that whenever a defendant's capacity for self-control does become a live issue in litigation, the entire context must be taken into consideration when establishing the standard of control that the law expects from the reasonable person. That is to say, the law must not hold the requisite standard of self-control expected from any defendant as immutable and unchanging. Rather, any such relevant factors that are fairly and reasonably recognised as depleting the self-control of any ordinary person and were operant at the time of an alleged offence *must* be taken into consideration. The failure to do so would be to ignore the reality that the capacity for self-control in reasonable people is something that ordinarily fluctuates, with empirical

evidence indicating the influence of a range of factors that are well-recognised for depleting self-regulatory resources.

A system of legal responsibility that does not apply such a contextualised approach to the question of self-regulatory capacities will inevitably hold people responsible for ordinary, every-day lapses in self-control that affect everybody in society. Equally, however, such a principle does not operate solely to diminish the degree of self-control that the law requires of ordinary people, as there are plenty of contexts where the hypothetical reasonable person may fairly be expected to pay extra attention to maintain a higher standard of control over their actions. One such obvious example concerns doctors, nurses and other medical staff, who must always deliver care to the standard of ordinary reasonable people *skilled in the same art*. The variability of self-regulation that ought reasonably to be expected by law is therefore bi-directional.

### *6.3.3. Meta-Control*

The present chapter has proposed that self-control is best understood as a capacity which requires mental resources and can be depleted over time. Moreover, in the moment of any given decision, it is this capacity for self-control which ultimately determines whether an individual proceeds to enact a particular (potentially criminal) decision into criminal action, or if they "exercise self-control" to veto that decision. In particular, the experiments concerning the unconscious veto considered in section 6.2, above, strongly suggest that the vetoing of any given decision or action is initiated without *necessarily* requiring conscious awareness or intervention. If this is correct, a fair question follows asking what role consciousness might play in self-control.

It is here proposed that consciousness permits for what is termed "meta-control", by which the capacity for automatic self-control can be modified and improved through exercise over time, much like a physical muscle. Thus, whereas the exercise of self-control in any particular moment may be understood as a function of the present capacity for automatic self-control and the mental resources available in order to sustain that capacity, that capacity for self-control itself can be gradually modified and improved over

time, for which purposes consciousness may play a role. Specifically, it may be consciousness that enables a person to both recognise the goal of improving their own self-control and recognise available strategies for achieving that goal, and effectively implement those strategies. For example, if a person is prone to anger quickly out of frustration, consciousness may be critical to enabling them to recognise the fact of their own diminished capacities for self-control and to practice strategies in order to improve this capacity.

An analogy might be drawn to the way in which a musician learns to play a new piece of music. Upon first sitting down to play an entirely new piece, the extent to which that person can immediately play that music by sight, both accurately and at speed, will be determined by their capabilities in that moment; the more accomplished musician will be more accurate and more able to play at speed. In this analogy, the individual's "talent" (*i.e.,* their ability to play a novel piece accurately and at speed) is a parallel for their capacity for automatic self-control. In order to improve their ability to perform a given piece of music, the musician adopts certain strategies: for example, they will concentrate more intently on the music that they are reading and their corresponding movements, they will slow down their playing, and they will approach the music in small sections, repeating each section and gradually increasing speed through a process of practice. This is a strategy that is consciously adopted in order to improve their performance of a particular piece of music.

Crucially, a significant improvement in the performance of a new piece of music does not emerge simply because a musician performs the aforementioned actions *in the moment*. Having applied this strategy once, for example, the musician may experience only a slight improvement in their ability to perform the piece of music accurately and at speed; simply practising the piece of music once does not produce a dramatic change to their capacity to perform it. Rather, it is the adoption and repetition of this strategy over time which produces the substantive change in the musician's ability to perform a new piece of music. By adopting a strategy of learning and practice, the musician improves their capacity to perform that piece accurately and at speed in any given moment. Self-control might be considered in a similar manner. Merely consciously concentrating on exercising self-

control in any given moment may be sufficient to deliver modest and relatively short improvements, but it is the repeated exercise of self-control over time which is instrumental to developing this capacity. Thus, it is here proposed that they key intervention that consciousness may ultimately play in the capacity for self-control is by enabling the recognition of the need (or goal) to improve self-control and, more pertinently, the adoption and practice of strategies *over time* which lead to the improvement of the capacity for automatic self-control *in the moment*.

## 6.4. From Self-Regulation and Control to Legal Responsibility

A number of general points may be drawn from the present chapter. First, the marshmallow test, related follow-up studies and other experiments investigating self-control, impulse control and delayed gratification, all point towards a highly determined component to these capacities and faculties. That is to say, a large degree of each individual's capacity for self-control in any given moment is strongly determined by a range of factors from their childhood development, upbringing and life experiences. Second, however, the strength model of self-control envisages this capacity akin to a muscle which can be exercised and trained over time.

The corollary of this follows that anybody's self-control can also be depleted; what is more, self-control depletion is highly correlated with criminal behaviour, with instances of psychiatric and personality disorders that impact upon self-control being overrepresented in prisons. Linking these two points is a possible vicious cycle that emerges, with poverty providing a factor which diminishes the development of self-control in childhood, diminished self-control being a significant factor that contributes to criminality, and criminality being a strong determinant of later poverty. Given the strongly determined component in the capacity for self-control, there emerges a large element of moral luck in whether or not any given individual is born into an environment which nurtures this capacity and other executive functions, or is born into the aforementioned vicious cycle.

Third, turning to the actual moment of any given action, the evidence considered in section 6.2 of this thesis suggests that the decision to veto a particular action arises first unconsciously in the brain, with conscious awareness thereof arising second. Indeed, in some of the studies reviewed, predictive brain signals indicating a subsequent veto decision could be recorded before subjects had even been presented with the initial decision that they would subsequently proceed to veto. These findings similarly follow those in relation to the *what*, *how*, and *when* components of a decision – all of the relevant components can and do operate automatically and without any *necessary* conscious intervention.

If one imagines a process of conscious deliberation – deciding to *x*, vetoing *x*, deciding to *y*, vetoing *y*, deciding to *z*, *etc.* – the research suggests that each stage of the deliberation – *i.e.,* each decision to do a thing, or to veto that thing and do another thing – is first reached unconsciously by the relevant automatic processes that are operating in parallel within the brain. Conscious awareness of each stage in that deliberation only arises second to the prior unconscious processes required to provide the answer to that stage. This, again, appears to introduce a large degree of moral luck in subsequent behaviour. Specifically, rather than the decision to do *x* or *y* or, indeed, to veto either, arising from necessarily conscious and deliberative processes, it is quite possible that a person decides to do *x* and decides not to veto that decision as a result of purely automatic and unconscious processes.

These two elements of moral luck present conceptual challenges to the current approach to *mens rea* within the topic of legal responsibility. The touchstone of responsibility is proof of subjective mental states – intention, recklessness, dishonesty, *etc.* – which coincide with the relevant prohibited criminal act. This is because people are presumed to act volitionally – that is, the capacities to exert conscious control over their decisions and actions, and to respond to good or bad reasons for so acting – such that a decision to commit a criminal act taken in the presence of the relevant guilty state of mind denotes moral wrongdoing, because the individual could have chosen to have done otherwise and controlled their actions accordingly. Whereas an individual defendant's volition is presumed to exist, refuting this presumption can form the basis of many criminal defences. At a more fundamental, philosophical level, the presumption of volition is underpinned

by the premise of free will – that people are free agents such that, when they freely choose to commit a criminal act, they may be punished. This latter philosophical perspective provides a theoretical basis for the current approach to *mens rea*, albeit is not a formal requisite component of the legal concept of responsibility.

The aforementioned elements of moral luck challenge these premises of *mens rea* in a number of ways. Considering the close correlation posited between self-control depletion and criminal behaviour, and the vicious cycle including poverty, the fact that many people may effectively be determined into criminality due to their underdeveloped capacity for self-control, challenges the premise upon which responsibility and punishment are justified because people decide and act freely. Whereas the various determinants are many, a combination of poverty, poor self-control, mental illness, *etc.,* inevitably coalesce into criminal conduct which, from a deterministic perspective, some individuals could scarcely ever have been expected to escape. At the same time, the law must nonetheless respond to criminal conduct arising from such determined self-control if it is to keep people safe and maintain peace and order.

Balancing these two contentions, it is incumbent upon the law have some measure of regard to the causes of defendants' diminished self-control specifically, and criminal conduct more generally, and give due recognition to causes of such conduct which lay outside of anybody's individual sphere of influence. One such way in which this might be achieved is through the introduction of a novel "not responsible" verdict, proposed in section 12.3.2 of this thesis, which would apply when an individual has committed the *actus reus* of an offence, but where they lacked the requisite *mens rea* due to the diminution of one or more mental capacities relevant to legal responsibility. Crucially, a verdict of not responsible would obviate the court's more punitive punishments, whilst still retaining the defendant under the jurisdiction of the court for the purposes of obtaining necessary medical treatment or other rehabilitation.

Considering the *whether* decision and the exercise of veto over a prepared plan of action, the second element of moral luck concerns whether or not a person's brain will, *in the moment of action*, produce the necessary activity in order to initiate a veto over that

planned action. As with the capacity for self-control generally, the exercise of self-control in a particular given moment appears dependent upon factors that lie firmly outside of an individual's conscious control and influence, namely the unconscious and automatic processes that are involved in vetoing a prepared action. Indeed, given that each of the *what, how, when* and *whether* components of a decision appear to operate automatically and without the necessary involvement of consciousness, there seems to arise a large degree of moral luck in the entire composite decision to do a given thing *x*.

That is to say, whether or not one individual over another arrives at a position of deciding to do a criminal act *x* and, crucially, *whether* or not they implement *x* into physical action, depends upon the happenstance of factors outside of an individual's direct conscious control – namely, the relevant automatic and unconscious brain processes arriving at the decision to do *x*. More pertinently, where the law presumes a high degree of direct, online conscious control over decisions and actions as part of the concept of volition, the evidence suggests that such decision-making and control over actions is largely automatic. At a more fundamental level, where the concept of *mens rea* is premised upon the notion that people are agents who make choices with free will, the neuroscience of decision-making reveals a mechanism of parallel processes operating automatically, unconsciously, and conceptually bound within an ultimately deterministic biological organism.

Fifth, it has previously been argued that the existence of subjective mental states such as intention or recklessness does not *in and of itself* provide proof that those mental states have been arrived at consciously and deliberately by a free agent, as opposed to arising automatically as the result of purely unconscious processing. A similar argument may be extended to the *whether* component of a decision – *i.e.,* the fact that an individual has proceeded to commit a certain (criminal) act is not *alone* proof that they consciously and deliberately chose to proceed with that act, as it is quite possible that their brain unconsciously and automatically decided not to veto the relevant actions.

The point is more clearly demonstrated in relation to offences that can be committed negligently, whereby negligence often denotes that the defendant has *omitted* to do something they were otherwise dutybound to do. In such circumstances, the fact that an

individual has (negligently) decided not to do *x* may arise because their brain unconsciously and automatically vetoed the decision to *x*, and not necessarily because the decision resulted from a conscious, deliberative and purposive process. As with the similar arguments presented in previous chapters, this renders reliance upon proof of subjective mental states as an ultimately unreliable, and potentially even arbitrary aspect of legal responsibility. The existence of a particular subjective state of mind does not alone provide proof that any free, conscious or deliberate choice to engage in criminal conduct has in fact been made.

Sixth, the discussion in the present chapter carries two related implications for theories of punishment, discussed in greater detail in chapter twelve of this thesis. On the one hand, the strongly determined component of the general capacity for self-control, alongside the automatic and unconscious operation of the decision component *whether* or not to commit a particular act, both undermine the premises of retributivism. Specifically, retributivism proposes that punishment is a moral good in itself, justified alone by the fact that people are agents who make free choices to commit moral wrongdoing in circumstances when they could decide to do otherwise. As has been discussed, the evidence suggests that self-control is a significantly determined capacity which operates moment-to-moment through automatic and unconscious processes. In this regard, whether or not an individual exercises self-control in any particular moment is a result of unconscious processes in the brain over which they have no direct conscious control. Consequently, in the specific moment when a person's acts, it is not the case that they could *consciously* decide to act otherwise.

On the other hand, the strength model of self-control discussed in section 6.3, above, provides the analogy of self-control as a muscle that can be trained and exercised gradually over time through practice. Considering the strong correlation between poor self-control and criminal behaviour, rehabilitative theories of punishment should be emphasised on the principle that, notwithstanding the strongly deterministic components of self-control, people can nonetheless be assisted with the tools and training necessary to improve upon those faculties of self-control. Notwithstanding that the exercise of self-control in any given moment appears to be the result of automatic and unconscious

processes, the improvement and strengthening of those processes in conjunction with the capacity to recognise and respond to good reasons for committing a criminal act or not, might fairly be expected to reduce future instances of criminal conduct which might otherwise have arisen as a result of poor or diminished self-control.

# 7. The *Why* Component, Access to Reason and Post-hoc Rationalisation

> 'Since an emotion is experienced as an immediate primary, but is, in fact, a complex, derivative sum, it permits men to practice one of the ugliest of psychological phenomena: *rationalization*. Rationalization is a cover-up, a process of providing one's emotion with a false identity, of giving them spurious explanations and justifications – in order to hide one's motives, not just from others, but primarily from oneself. The price of rationalizing is the hampering, the distortion and, ultimately, the destruction of one's cognitive faculty. Rationalization is a process not of perceiving reality, but of attempting to make reality fit one's emotions.'
>
> - Ayn Rand, 1982.[1]

Writing in 1977, Nisbett and Wilson offer a comprehensive discussion of dozens of studies concerning the subjective access to mental processes, as well as reporting the results of a number of their own smaller studies. They concluded that 'subjects are sometimes (a) unaware of the existence of a stimulus that importantly influenced a response, (b) unaware of the existence of the response, and (c) unaware that the stimulus has affected the response', such that 'when people attempt to report on their cognitive processes… they do not do so on the basis of any true introspection.'[2] From their own studies, the authors found that subjects were 'virtually never accurate in their reports' of the external or internal stimuli influencing their decision-making, which largely corroborates the findings from other research.[3] For the purposes of this discussion, a given reason may be regarded as a *genuine* reason for doing a particular thing (as contrasted with one that is false, confabulated, or otherwise insufficient) 'if it is the case

[1] Ayn Rand, *Philosophy: Who Needs It* (Signet 1982), 18.
[2] Richard E. Nisbett and Timothy DeCamp Wilson, 'Telling more than we can know: Verbal reports on mental processes' (1977) 84(3) *Psychological Review* 231, 231.
[3] *Ibid*., 242 – 243.

that one is prepared not to do it, or to stop doing it, should those reasons be found not to be reasons for doing it and there are no other reasons for doing it.'[4]

One such example in Nisbett and Wilson's research concerned unconscious priming, discussed further in section 7.1.3, below. Subjects were instructed to memorise word pairs which were intended to generate associations that would be elicited later in a word-association task. This produced an average effect of doubling the frequency that target responses would be returned in the word-association exercise. Thus, subjects primed by memorising the pair "ocean-moon" were more likely to name a target laundry detergent, "Tide." However, despite most of the subjects being able to recall the majority of word pairs, they 'almost never mentioned' that cue as a reason for selecting the target detergent; more interestingly, they proceeded to recite some distinctive feature, personal meaning, or affective reaction to the detergent which they named.[5] This disconnection between subjectively reported and genuine reasons for decisions not only operates where subjects report stimuli as being influential on their decision when it is in fact not so, but similarly where subjects fail to identify stimuli which was indeed demonstrably influential.

Nisbett and Wilson are keen to highlight that the various studies reviewed 'do not suffice to show that people *could never* be accurate' about the subjective reasons for, and processes behind, particular decisions,[6] and the same caveat applies to the further studies considered in this chapter. That notwithstanding, the various studies do strongly indicate that 'such introspective access as may exist is not sufficient to *produce accurate reports* about the role of critical stimuli.'[7] They therefore suggest that, *sometimes* people are unable to correctly report the existence of evaluative and motivational responses to manipulations, the occurrence of a cognitive process, or the existence of critical stimuli, whilst 'even when people are completely cognizant of the existence of both stimulus and response, they appear to be unable to report correctly about the effect of the stimulus on the response.'[8] Whereas Nisbett and Wilson's conclusions are correctly posed as

[4] Kevin Magill, 'The idea of justification for punishment' (1998) 1(1) *Critical Review of International Social and Political Philosophy* 86, 90.
[5] Nisbett and Wilson (1977), 243.
[6] *Ibid*., 246.
[7] *Ibid*., 246.
[8] *Ibid*., 246 – 247.

occurring "sometimes", a critical issue exists in the near impossibility for individuals to subjectively discern which reasons for action are genuine and which are not.

This chapter proceeds to consider some of the more modern experiments which cast further doubt on the ability of individuals to subjectively access the reasons for decision-making; that is the stimuli contributing to, and cognitive process involved in, reaching a particular decision. It is suggested that, even if individuals do have a greater degree of conscious control over decision-making than has been suggested in the preceding chapter, they nevertheless have a diminished ability to accurately recall the reasons for particular decisions and, furthermore, a significantly diminished ability (if any) to distinguish between accurate and inaccurate reasons for decisions. Consequently, it remains irrational for legal responsibility to rest on such subjective states of mind as intention, recklessness or dishonesty, *etc.*, the existence of which cannot reliably be recalled by individuals reporting on their subjective mindset, nor definitively observed from an objective viewpoint. Put differently, how can the law differentiate the attribution of legal responsibility between a killer who intended the death of their victim (murder) and one who did not (manslaughter), when neither can be relied upon to accurately report the causes of their decision or cognitive processes which resulted therein, and nor can observers necessarily objectively distinguish the two.

## *7.1. Subjective Access to and Production of Reasons*

### *7.1.1. The Split-Brain Experiments*

Roger Sperry, Joseph Bogan and Michael Gazzaniga conducted some of the earliest pioneering experiments which explored the activities of the two cerebral hemispheres, observing through subjects for whom the hemispheres had been surgically separated for therapeutic purposes.[9] The corpus callosum is the structure that connects the two hemispheres of the brain, the severing of which was infrequently used in the treatment of

[9] See Michael S. Gazzaniga, Joseph E. Bogen and Roger W. Sperry, 'Some functional effects of sectioning the cerebral commissures in man' (1962) 48(10) *Proceedings of the National Academy of Sciences* 1765; Michael S. Gazzaniga, Joseph E. Bogen and Roger W. Sperry, 'Observations on visual perception after disconnexion of the cerebral hemispheres in man' (1965) 88(2) *Brain: A Journal of Neurology* 221; Michael S. Gazzaniga and Roger W. Sperry, 'Language after section of the cerebral commissures' (1967) *Brain: A Journal of Neurology* 131.

severe cases of epilepsy. With each hemisphere receiving sensory information from contralateral sides of the body but no longer able to communicate with one another, it was possible for experimenters to effectively communicate with the two hemispheres individually, thereby controlling the information available to one or both and observing its effects through features dominant to each hemisphere, such as language, speech, writing and motor tasks. The paradigm setup involved separating the subject's field of vision in half, such that items could be shown to the right-brain via the left eye and *vice versa*, but each hemisphere did not have access to the other's stimuli. As Bennett and Hacker summarise, split-brain patients could verbally describe an object presented in the right visual field, the left hemisphere exhibiting dominance for language and speech. If the object was presented in the left visual field, however, the subject could not say what it was, but could point to a similar object. They note, 'similar results were found for the other sensory modalities of touch, sound and smell.'[10]

A decade later, LeDoux and Gazzaniga reported further trials with a subject whose right brain, somewhat uniquely, had linguistic abilities permitting responses to verbal commands, albeit not through speech.[11] A first example occurred where the right hemisphere (lacking speech) was instructed to laugh, and the subject responded accordingly. However, when the subject was asked why he had laughed, the left hemisphere provided the explanation "Oh you guys are really something."[12] In a second example, the subject was provided with simultaneous images of a snow scene to the right hemisphere and a chicken claw to the left, after which each hemisphere was presented with pictures and instructed to select the one most relevant to the image it had seen. The right hemisphere selected a shovel associating with the snow scene, and the left hemisphere chose a picture of a chicken, matching the claw. Interestingly, however, when asked to explain the choices, the left hemisphere (possessing speech) explained, "I saw a claw and I picked a chicken, and you have to clean out the chicken shed with a shovel."[13]

[10] Maxwell R. Bennett and Peter M. S. Hacker, *History of Cognitive Neuroscience* (Wiley-Blackwell 2008), 11.
[11] See Joseph E. LeDoux, Donald H. Wilson and Michael S. Gazzaniga, 'A divided mind: Observations on the conscious properties of the separated hemispheres' (1977) 2(5) *Annals of Neurology* 417; Michael S. Gazzaniga and Joseph E. LeDoux, *The Integrated Mind* (Plenum Press 1978).
[12] Gazzaniga and LeDoux (1978), 146.
[13] *Ibid*., 148.

In other examples, the subject's right hemisphere had responded to instructions to stand up or rub their head which, when asked to explain why they had performed that action, the left hemisphere again invented explanations in the absence of the information available only to the right.

As Klein and Kihlstrom explain, in each of the examples the left hemisphere was presented with the same challenge:

> '[I]t had observed a response but did not know why the response was performed. When asked "Why are you doing that?", the talking left hemisphere had to come up with a plausible explanation for a behaviour performed in response to a command directed to the mute right hemisphere.'[14]

The left hemisphere attempts to fill this gap by providing credible reasons for actions and behaviour. Thus, as LeDoux, Wilson and Gazzaniga write, the 'conscious verbal self' is not always aware of the underlying reasons or motivations for doing a particular action and, when tasked with explaining that action, 'it attributes cause to the action as if it knows, but in fact it does not.'[15] Crucially, the verbal left hemisphere does not 'offer its suggestion in a guessing vein, but rather as a statement of fact' as to why that action was performed.[16]

LeDoux and Gazzaniga consider it to be likely that this phenomenon would also occur in neurotypical patients whose left and right hemispheres were still connected; they denoted the left hemisphere as containing an "interpreter" function 'whose job it is to interpret our behaviour and our responses, whether cognitive or emotional, to environmental challenges.'[17] However, this interpreter can only be 'as good as the information it

[14] Stanley B. Klein and John F. Kihlstrom, 'On bridging the gap between social-personality psychology and neuropsychology' in Cacioppo J. T. and Berntson G. G. (eds.), *Foundations in Social Neuroscience* (Massachusettes Institute of Technology 2002), 55.
[15] Joseph E. LeDoux, Donald H. Wilson and Michael S. Gazzaniga, 'Beyond commissurotomy: Clues to consciousness' in Gazzaniga M. S. (ed.), *Handbook of Behavioral Neurobiology Volume 2: Neuropsychology* (Plenum Press 1979), 549.
[16] *Ibid*.
[17] Michael S. Gazzaniga, *The Mind's Past* (University of California Press 1998), 174.

receives.'[18] Consequently, in circumstances where, for whatever reason, the interpreter lacks access to the reasons or motivations for an action, it is adept a presenting the best plausible explanation which the subject experiences as being the factual cause of their behaviour. It has since been suggested that the right hemisphere will also confabulate its interpretation of events when the requisite information concerning motives is unavailable,[19] and Hirstein summarises, 'whenever a situation is set up so that the two hemispheres can be assessed for confabulation on an equal basis, the right hemisphere registers as equally confabulatory.'[20]

The key point being extrapolated from the split-brain experiments is that the brain both *can* and *does* provide reasons for actions which do not accurately explain the actual motives or causes of those actions. Moreover, these reasons are not offered as suggestions, possibilities or deceptions, but are provided by the brain as genuine reasons – they are confabulations without the intention to deceive. As Hirstein continues to summarise, sceptics might claim that we lack access to our reasons or intentions most, if not all of the time; however, even those on the opposing side do not tend to suggest that we always know our intentions, but rather 'that we *can* know what they are, especially in the case of nonroutine, more cognitively challenging actions.'[21] It is of particular note that these confabulations would generally – and, at least, not without further contemplation – fall into the category of being unknown, *i.e.,* we are not generally aware of which reasons correctly relate to genuine causes of our actions, and which reasons are confabulated by the brain. Thus, whether or not the conscious brain has accurate access to the true causes of decisions, this is of quite limited value if it remains practically impossible to discern *which* consciously recalled reasons for action are correct, and which are confabulations.

[18] *Ibid*., 136.
[19] For example, see Lei H. Lu, Anna M. Barrett, Ronald L. Schwartz, Jean E. Cibula, Robin L. Gilmore, Basim M. Uthman and Kenneth M. Heilman, 'Anosognosia and confabulation during the Wada test' (1997) 49(5) *Neurology* 1316.
[20] William Hirstein, *Brain Fiction: Self-deception and the Riddle of Confabulation* (Massachusetts Institute of Technology 2005), 164 – 166.
[21] *Ibid*., 176.

### *7.1.2. Transcranial Magnetic Stimulation*

Brasil-Neto *et. al.* investigated the effects of transcranial magnetic stimulation ('TMS') on a 'warned, forced-choice response time task' performed by normal adult subjects.[22] TMS is a non-invasive neurostimulation procedure which uses magnetic fields to stimulate electrical activity in small specifically targeted regions of the brain. Subjects were instructed to extend their index finger in response to a go signal which was the "click" of the TMS device being activated. The subjects could choose when, after the go signal, to extend their finger and which finger to use; however, they were instructed to make this choice after the go signal was produced and, crucially, they were unaware in each instance whether or not the TMS device was pointed towards or away from their head.

When TMS was applied to the motor area of the brain, subjects were more likely to choose the hand contralateral to the hemisphere stimulated, typically responding in less than 200 milliseconds following the go signal; in longer response times the magnetic stimulation had no effect on hand preference.[23] Brasil-Neto *et. al.* theorised that the magnetic stimulus could be interacting with mechanisms for disinhibition or directly activating response channels in the brain – 'it could either reduce the threshold of response channel activation or increase the rate of activity build up in the response channel, causing a fixed threshold to be reached earlier.'[24] Crucially, however, the subjects remained unaware of whether or not the TMS had influenced their choice of finger, demonstrating the possibility for influencing movement preparation in the brain 'without disrupting the conscious perception of volition.'[25]

Brasil-Neto *et. al.* link this work to Libet's studies, discussed above in the chapter five, in which cerebral activity was found to precede a conscious intention to act by at least 200 milliseconds, commenting that their results concur with Libet's findings and suggest that:

---

[22] Joaquim P. Brasil-Neto, Alvaro Pascaul-Leone, Josep Valls-Solé, Leonardo G. Cohen and Mark Hallett, 'Focal transcranial magnetic stimulation and response bias in a forced-choice task' (1992) 55(10) *Journal of Neurology, Neurosurgery, and Psychiatry* 964, 964.
[23] *Ibid*., 964 – 965.
[24] *Ibid*., 965.
[25] *Ibid*.

> '[C]onscious perception of willing a particular action is preceded, and possibly generated, by cerebral processes that can be influenced by magnetic stimulation. Since conscious perception and the resulting movement can be consistently and predictably influenced by magnetic stimulation of the motor areas, these early cerebral processes probably account for the generation of both the conscious perception of wanting to move and the corresponding movement.'[26]

For the purposes of this chapter, and linking the study by Brasil-Neto *et. al.* to the split-brain experiments, the key point is that subjects were unaware when their choice of finger was influenced by TMS; their reason for making one choice or the other in any instance could not be given by reference to one of the actual key causes, this being the magnetic stimulation.[27] Brasil-Neto *et. al.* may also be linked with the 1991 study by Fried *et. al.* in which electrical stimulation was applied directly to the brain during surgery, discussed in sections 3.2 and 5.2.4, above. Both studies note how the sensation of an intention to act can be influenced through magnetic and electrical stimulation respectively. Linking both to the split-brain experiments is the question of whether subjects could discern between genuine reasons for action and reasons being generated without volition, *i.e.,* through external stimulation. In the aforementioned study by Fried *et. al.* this question was not investigated, whilst in the split-brain experiments and the studies from Brasil-Neto *et. al.*, the subjects' reasons for particular choices were not concurrent with the actual stimulation, this being respectively the images transmitted to the right hemisphere or TMS applied to the brain.

In a study similar to that conducted by Brasil-Neto *et. al.*, discussed above, subjects were instructed to choose between extending their right or left index finger after hearing a brief "click" stimulus.[28] This stimulus was caused by a magnetic coil delivering TMS the intensity of which was below the threshold for eliciting a motor response. The coil was

[26] *Ibid*., 966.
[27] See further Daniel M. Wegner, 'The mind's best trick: how we experience conscious will' (2003) 7(2) *Trends in Cognitive Sciences* 65.
[28] Klaus von Ammon and Simon C. Gandevia, 'Transcranial magnetic stimulation can influence the selection of motor programmes' (1990) 53(8) *Journal of Neurology, Neurosurgery, and Psychiatry* 705.

clamped in place over various brain regions and delivered a single magnetic stimulus for each repeated trial, investigating whether this subthreshold intensity for motor response could nevertheless influence the selection by subjects of one hand or the other. The study found a 'highly significant' correlation between the direction of the TMS current applied and the hand selected by subjects in each trial; 'when questioned they felt that their decisions appeared to be made in an entirely natural way.'[29]

The principal finding extrapolated by Ammon and Gandevia is that 'cortical processing can be influenced by levels of magnetic stimulation which are "subthreshold" for a motor response.'[30] The significance for the purposes of the present discussion, however, is similar to that found in the TMS study by Brasil-Neto *et. al.*, namely the point concerning subjects' *inability* to discern the primary cause of their choices, *i.e.,* the TMS. This further demonstrates the apparent separation between the actual reasons for (or causes of) decisions on the one hand, and those subjectively accessible or reportable by subjects on the other. Interpreting the TMS studies by Brasil-Neto *et. al.* and Ammon and Gandevia, alongside the seminal paradigm by Libet *et. al*. and its repetition by Haggard and Eimer, Haggard writes that 'all these results suggest the interesting possibility that the process of selecting between alternative actions, which philosophers often consider the core of "free will", could result from routine processes *operating automatically and unconsciously*.'[31]

### *7.1.3. Unconscious Priming*

The phenomenon of priming and its automatic and unconscious operation has been further discussed substantively in chapter three of this thesis. Chartrand and Bargh[32] describe two experiments supporting the hypothesis that 'the effect of activated goals is the same whether the activation is nonconscious or through an act of will.'[33] The first experiment

[29] *Ibid*., 706.
[30] *Ibid*.
[31] Patrick Haggard, 'Conscious intention and motor cognition' (2005) 9(6) *Trends in Cognitive Sciences* 290, 292 (emphasis added).
[32] Tanya L. Chartrand and John A. Bargh, 'Automatic activation of impression formation and memorization goals: Nonconscious goal priming reproduces effect of explicit task instructions' (1996) 71(3) *Journal of Personality and Social Psychology* 464.
[33] *Ibid*., 464.

replicates much of the methodology and results from an earlier paradigm;[34] subjects read a series of sentences describing the behaviour of a target person, and were unconsciously primed to either form an impression of the target described, or to memorise as much information as possible (in the original experiment subjects were instructed explicitly to either perform the impression-forming or memorisation task). Chartrand and Bargh's results mirrored those from Hamilton, Katz and Leirer's earlier paradigm, showing a significant effect of priming with the clustering and recall of target information; 'the information-processing goals that have been shown in previous work to produce differential organization and memory for social information when operating consciously and intentionally *have the identical effects on processing when operating automatically*.'[35]

The second experiment also replicates earlier paradigms;[36] subjects were presented with a series of behaviours which were either consistent or inconsistent with a particular personality trait, such as honesty. However, the subjects were not explicitly instructed to perform an impression-forming task but, rather, one group was subliminally primed with words corresponding to impression formation whilst a second group was exposed to words unrelated to forming impressions. The experiment demonstrated that the extent of impressions formed on-line – *i.e.,* impressions formed at the time that relevant information is presented as opposed to later in time – 'differed reliably as a function of whether the impression formation goal had been primed subliminally.'[37]

These two experiments replicated previous social cognition paradigms with the exception that processing goals were subliminally primed as opposed to being explicitly instructed. In replicating the findings of earlier experiments, the work by Chartrand and Bargh 'strongly support' the proposition that 'intentions and goals can be automated and that

[34] David L. Hamilton, Lawrence B. Katz and Von O. Leirer, 'Cognitive representation of personality impressions: Organizational processes in first impression formation' (1980) 39(6) *Journal of Personality and Social Psychology* 1050.
[35] Chartrand and Bargh (1996), 469.
[36] Reid Hastie and Purohit A. Kumar, 'Person memory: Personality traits as organizing principles in memory for behaviours' (1979) 37(1) *Journal of Personality and Social Psychology* 25; and John A. Bargh and Roman D. Thein, 'Individual construct accessibility, person memory, and the recall-judgment link: The case of information overload' (1985) 49(5) *Journal of Personality and Social Psychology* 1129.
[37] Chartrand and Bargh (1996), 472.

their effects when operating non-consciously are identical to their effects when they are operating consciously and deliberately.'[38] Interpreted at its extreme, this work might suggest that there is little or no difference between intentions and goals which are formed consciously and those which operate nonconsciously, save perhaps for the fact of there being conscious awareness of those goals in action. This extreme view would correspond with the limited (or non-existent) role of conscious control in decision-making implied in previous chapters of this dissertation and following the work of Libet *et. al.* However, such an extreme position is not explicitly demonstrated in these experiments.

Alternatively, a narrow interpretation is concurrent with the same problem raised by the split-brain experiments; in the same way that split-brain subjects were unable to distinguish between genuine reasons for decisions and confabulations provided by the brain, the difference between goals which are adopted and effected automatically and those which are conscious and explicit may be similarly indistinguishable. This raises issues where the law attempts to attribute responsibility on the basis of explicitly held goals (or intentions). In one example, a defendant may have unconsciously determined to follow a particular criminal course of action, which unconsciously affects their behaviour towards realising that goal; in a second example, a defendant consciously determines to commit a crime and proceeds to do so. This second example describes a "neuro-typical" adoption and progression of a particular intention which would in turn attract legal responsibility. The first example, however, could easily describe the actions of somebody behaving as an automaton who has neither conscious awareness nor control over a particular intention. The law distinguishes between these two positions, whereas the experiments described by Chartrand and Bargh suggest that people may not in fact be able to do so.

*

Bargh, Lee-Chai, Barndollar, Gollwitzer and Trötschel describe a number of experiments demonstrating the proposal that 'goals can be activated outside of awareness and then

[38] *Ibid.*, 475.

operate nonconsciously to guide self-regulation effectively,' [39] building upon the experiments described above by Chartrand and Bargh. In the first experiment, subjects were assigned to either a high-performance goal or neutral priming condition implemented through a word-search puzzle. Each puzzle contained six neutral words and seven words either relevant or not to the ideal of high performance. These priming puzzles were followed by three experimental puzzles, each with hidden words relating to a theme; in post-experimental interviews, no subject reported any awareness or suspicion as to the relationship between the first priming puzzle and the subsequent experimental puzzles. Those subjects which were primed with the goal of performing well exhibited an enhanced performance of the experimental puzzles; this demonstrates that 'performance goals can become activated without the necessity of conscious and deliberate choice and then operate to regulate behaviour towards attainment of the desired outcome.'[40]

The second experiment presented subjects with a resource-dilemma task in which they would play with a presumed other participant, harvesting resources from a pool which would be periodically replenished. The game enabled subjects to adopt a competitive strategy accruing higher "profits" but exhausting the pool more quickly, a cooperative strategy in which profits were returned to the common pool, or attempting a strategy which achieved both profit and common good. The design permitted a comparison of both subjects who were and were not subliminally primed with a high-performance goal, and further between subjects whose primed goal operation was concurrently consciously held against those who were primed without any conscious awareness of the goal.

Both priming and conscious goal-setting for cooperation were found to produce corresponding behaviour, with greater cooperation amongst participants given an explicit conscious goal to do so. More importantly, however, a similar effect was shown between subjects primed and given the conscious goal, 'providing a second demonstration of non-conscious goal activation along a different dimension of behaviour' than in the first

[39] John A. Bargh, Annette Lee-Chai, Kimberly Barndollar, Peter M. Gollwitzer and Roman Trötschel, 'The automated will: Nonconscious activation and pursuit of behavioral goals' (2001) 81(6) *Journal of Personality and Social Psychology* 1014, 1014.
[40] *Ibid*., 1021.

reported experiment.[41] Moreover, subjects who were not provided with a conscious goal 'showed the same increase in cooperation due to goal priming as did participants in the conscious-goal condition', confirming the authors' hypothesis that nonconscious goal activation *does not* require the 'preexistence of a conscious goal in the same direction' – there was 'no association between consciously experienced strength of goal intention and the actual effect of goal priming on behaviour.'[42]

A third experiment sought to confirm whether or not the effects from the previous two were the result conscious or unconscious goal activation, concluding in favour of the latter. In the fourth experiment, Bargh *et. al.* confirmed the hypothesis that pursuit of a nonconsciously primed goal would share in critical features of the pursuit of consciously selected goals, namely that individuals would persist on a given task in spite of obstacles. The fifth and final experiment explored whether subjects resumed their pursuit of an unconsciously activated goal that had been interrupted for five minutes, finding indeed that 'participants with a nonconscious goal to attain high performance were considerably more likely to return to the incomplete intellectual task after the interruption than were nonprime participants.'[43]

Interpreting the findings of each experiment together, Bargh *et. al.* conclude that behavioural goals may be activated without the requirement for any conscious decision or even awareness over that goal. In turn, nonconscious goals proceed to operate similarly to consciously adopted goals – 'they promote goal-directed action…, they increase in strength until acted on…, they produce persistence at task performance in the face of obstacles…, and they favor resumption of disrupted tasks even in the presence of more attractive alternatives.'[44] Furthermore, the second experiment in particular demonstrated the effects of such priming absent of any concurrent conscious awareness of the goal being primed. In subject interviews conducted after the fact, the reported intentions of subjects primed to be cooperative were unrelated to the degree of cooperation actually exhibited, even though the priming had produced a cooperation-goal effect in their

[41] *Ibid.*, 1023.
[42] *Ibid.*, 1023 – 1024.
[43] *Ibid.*, 1024.
[44] *Ibid.*, 1033.

subsequent behaviour. This suggests not only that nonconscious priming does not depend on any conscious awareness but, furthermore, supports the notion that 'participants who are unaware of the activation of nonconscious goals will remain unaware of their subsequent operation to guide behaviour.'[45]

Indeed, following a comprehensive review of dozens of priming studies up to 2010, Smeester, Wheeler and Kay observe that 'in all the prime-to-behavior-studies to date, the prime recipients remain unaware that the prime played any role in their behavior. That is, in the debriefing of participants in these experiments, participants are unable to identify the priming task as having any effect on their behaviour.'[46] This, again, raises crucial questions regarding the *why* component of decision-making, of whether and to what extent people are able to accurately access genuine reasons for their decisions and actions. In particular, the current chapter and chapter three of this thesis each present a broad range of priming studies, demonstrating the subtle yet influential effects that priming can exert over decision-making processes. If people are indeed rarely (if ever) able to distinguish goals, motivations and intentions that have been generated endogenously from those that have been externally primed from some exogenous source, this calls further into question the ability for people to accurately and reliably access genuine subjective reasons for their decisions.

### *7.1.4. The Legal Relevance of Split-Brain Experiments, TMS and Unconscious Priming*

In the process of prosecuting any given offence, the court will enquire not only as to what people did, but also why they did so, in order to establish or, indeed, negate the existence of the relevant *mens rea*. As with the *what, how, when,* and *whether* components of a decision, the evidence considered in the present section suggests that the *why* behind a decision – that is, the *genuine* reason for action as defined in the introduction to the present chapter – may also exist and operate in the unconscious brain without any

[45] *Ibid*., 1034.
[46] Dirk Smeesters, S. Christian Wheeler and Aaron C. Kay, 'Indirect prime-to-behavior effects: The role of perceptions of the self, others, and situations in connected primed constructs to social behavior' in Zanna M. P. (ed.), *Advances in Experimental Social Psychology: Volume 42* (Elsevier Academic Press 2010), 307.

necessary concurrent conscious awareness. What is more, the evidence from priming studies shows how the reasons behind actions, like action goals and intentions, may be primed within a person from exogenous sources, and proceed to influence their mental processing without any necessary conscious awareness of intervention.

Further still, considering the research in sections 7.1.1 and 7.1.2. of this chapter, the evidence suggests that people have relatively poor abilities to distinguish between genuine reasons for action and confabulations created by the brain to "fill in the gaps" in its knowledge and interpretation of events, *i.e.,* gaps in the information it possesses. In this regard, the court's investigation into *why* a defendant acted as they did cannot *necessarily* provide any insight into a person's *genuine* reasons for action, and may instead merely serve to test how convincingly their brain can confabulate. Put differently, once again, the focus on subjective mental states of mind as a key determinant of responsibility may be inherently unreliable, as people are demonstrably poor at introspecting genuine from confabulated reasons, and nor is it any easier for observers to objectively distinguish between the two. These propositions are expanded upon further in the following sections of this chapter.

## *7.2. Justifying Moral Decisions*

A number of further experiments in psychology explore moral reasoning and the justifications people give for moral judgments. This is particularly interesting within the context of the present thesis given the intrinsic similarities between legal and moral dilemmas regarding which people must deliberate and make decisions. What many of these experiments demonstrate – akin to the split-brain studies – is that people appear to have a relatively poor ability to provide sound or reasonable justifications for their moral judgments. It is submitted that this further supports the central propositions of this chapter, that we generally have an unreliable access to the true reasons for many of our decisions and, moreover, that there is little or no ability to accurately distinguish between those reasons which are genuine, confabulated, or unconsciously primed by exogenous sources.

Looking at moral judgments and decision-making is particularly insightful to this thesis for a number of reasons. First, the kinds of moral problems used and the time available for subjects to consider their responses allows a degree of insight into a more consciously deliberative thought process, in contrast to some of the faster and less meaningful decisions that have formed the basis of other studies, such as pressing buttons. Moreover, there is an undeniable congruence between the legal and moral aspects of responsibility; the legal proscription against murder or theft, ideas of intent or dishonesty, and even normative principles such as double effect and the distinctions between actions and omission, all appear prominently in both legal and moral discourse. The experiments concerning people's moral reasoning therefore provide a window into how some of the moral dilemmas underlying many legal problems are resolved by ordinary people (*i.e.,* non-lawyers).

Hauser, Cushman, Young, Jin and Mickhail report a fascinating study exploring the dissociation between people's moral judgments and the reasons offered in justification.[47] Subjects were visitors to a test website which attracted around 5,000 responses from people in 120 different countries (albeit with a strong bias towards English-speaking nations). The subjects were presented with the familiar philosophical "trolley problem" thought experiment, in which a runaway train will hit five people on the track unless the subject chooses to divert the train to an alternative track, hitting only one person instead. In brief, four scenarios were presented: scenarios 1 and 2 involved the choice between turning the train by steering it to a side track or pushing a large man onto the tracks, exploring the use of the principle of the double effect,[48] the relevance of physical contact, and the introduction of a new threat contrasted with redirecting an existing threat. Scenarios 3 and 4 explored the double effect principle in closer detail; in both instances the train was diverted impersonally by throwing a switch, whilst in scenario 3 the object blocking the train was the large man (he was an intended means of saving others), whereas

[47] Marc Hauser, Fiery Cushman, Liane Young, R. Kang-Xing Jin and John Mikhail, 'A dissociation between moral judgments and justifications' (2007) 22(1) *Mind & Language* 1.

[48] The principle of double of effect refers to the argument that causing harm is permissible when it is the side effect (or "double effect") of bringing about some intended good, and that harm is neither itself intended nor a necessary means of achieving the intended good.

in scenario 4 the man stood before another large object stopping the train, and thus his death was a foreseen but unintended side effect.[49]

Subjects first reported whether or not they would choose to divert the train by sacrificing one person in order to save five, after which they were invited to explain their reasons and, in particular, to account for why scenarios 1 and 2 or 3 and 4 were judged differently when the outcomes were identical. The responses were ordered into one of three categories: a "sufficient" justification, interpreted widely as one which 'correctly identified any factual difference between the two scenarios and claimed the difference to be the basis of moral judgment.'[50] These typically included reference to the necessity of one man's death in saving five, redirecting an existing threat versus introducing a new one, and actions which were impersonal as opposed to personal or 'emotionally salient.'[51]

The second category was an "insufficient" justification which failed to identify any factual difference between the scenarios. These responses typically fell within three types: an inability to account for their contrasting judgments between the scenarios, appealing instead to "reasonableness" or gut feeling; referring to death or killing as 'inevitable' in one scenario but not another whilst offering no further explanation behind this reasoning; or using utilitarian reasoning in one scenario and deontological in another, but 'without resolving their conflicting responses.'[52] The third category of responses were either blank or included new assumptions, such as "people working on the track are responsible whilst those walking alongside it are reckless."[53] The third category responses were excluded for being based on assumptions which could not be drawn from the scenarios presented. However, Hauser *et. al.* cite Cushman *et. al.* (below)[54] in postulating that these responses are likely to be confabulations which arise 'because subjects are incapable of accounting for the pattern of their judgments.'[55]

---

[49] *Ibid*., 6.
[50] *Ibid*., 13.
[51] *Ibid*.
[52] *Ibid*., 14.
[53] Paraphrased.
[54] Fiery Cushman, Liane Young and Marc Hauser, 'The role of conscious reasoning and intuition in moral judgment' (2006) 17(12) *Psychological Science* 1082.
[55] Hauser *et. al.* (2007), 14.

A significant majority (85%) of responses considered impersonally throwing the switch and causing foreseeable death to be morally permissible, whilst an equally significant minority (12%) considered that personally pushing a man onto the tracks causing intentional death was permissible. These differences in responses could be related to the principle of double effect, the fact of pushing another individual personally, or the introduction a new threat as opposed to diverting an existing one. In the scenarios which further explore the double effect principle, 56% of subjects considered that diverting the train and causing intentional harm was permissible, whilst 72% believed doing so and causing foreseeable (but not intentional or necessary) harm to be acceptable. This statistically significant difference indicates that, 'as a group, subjects make use of the principle of the double effect.'[56]

Hauser *et. al.* tested subpopulations for which they had collected sufficient statistical power to detect significant discrepancies. Across almost all subpopulations including age, gender, level of formal education, and formal exposure to moral philosophy, the results remained consistent. There was a small but statistically significant difference in the proportions of Catholics, Protestants and Atheists who judged the permissibility of scenarios 3 and 4 differently but, interestingly for a study of this nature, 'there were no significant differences between subjects who had and had not taken formal coursework on moral philosophy.'[57] This consistency of results was also found in relation to the ability to provide sufficient justifications for actions.

Across all subpopulations (including, perhaps surprisingly, subjects with some formal education in moral philosophy) 44% of responses had to be excluded for falling into third category and making unsupported assumptions. Of the remaining responses, 70% provided insufficient justifications for differences in their moral judgments, and only 30% could offer sufficient reasons. Of this 30%, the moral philosophers did have a statistically significant advantage in offering a sufficient justification for their decisions.[58] The authors expanded the data sets regarding scenarios 3 and 4, exploring the double effect

[56] *Ibid.*, 8.
[57] *Ibid.*, 11.
[58] *Ibid.*, 15.

principle. They found that 33% of respondents used the principle in judging the scenarios differently, with foreseeable harm regarded as being permissible but intentional harm as impermissible. In the reasons given, two-thirds had to be excluded for making unsupported assumptions; only 2% provided a sufficient reason.

Between this additional data-set exploring the double effect principle in particular and the main trial, the results suggest respectively, with a 95% level of confidence, that only between 2% and 34% of individuals judging the scenes differently 'would be able to provide a sufficient justification for their judgments.'[59] Two particular findings are extrapolated by the authors:

> '[I]n the context of the trolley problems we studied, all of the demographically defined groups tested within our sample showed the same pattern of judgments and *subjects generally failed to provide justifications that could account for the pattern of their judgments*.'[60]

Hauser *et. al.* continue to submit that these findings raise challenges to the view that moral judgments are significantly produced as a result of 'conscious reasoning from a set of moral principles.'[61] If this were the case, they argue, it would be expected that subjects educated in moral philosophy should at least be more likely to invoke the double effect principle, or some other relevant argument, than other subjects who lacked such exposure to philosophical discourse. More generally, the conscious reasoning perspective might also expect differences in beliefs and attitudes according to other demographic characteristics; however, 'at least for the principles tested, and at least within the range of variation of our subject population, the conscious reasoning perspective cannot account or the pattern of results.'[62]

The authors equally note that those subjects who judged scenarios 3 and 4 to be different, and could therefore only be applying the principle of double effect, also generally failed

[59] *Ibid*., 14 – 15.
[60] *Ibid*., 15 (emphasis added).
[61] *Ibid*., 16.
[62] *Ibid*., 16.

to appeal to this principle or describe its central distinction between intended means and foreseeable side effects. This, again, points away from moral decision-making resulting from conscious reasoning built upon sets of fundamental moral principles.[63] None of this is to say that conscious reasoning cannot be used when addressing such moral alternatives; but that this is unlikely to be the dominant mode of moral decision-making, which may be altogether more intuitively and emotionally mediated. The study also demonstrates a further point more pertinent to this chapter of the thesis; even though we 'sometimes deliver moral judgments based on consciously accessed principles, often we fail to account for our judgments.'[64] Thus, the research further suggests towards an unreliability in our powers of consciously introspecting and recounting genuine reasons for decisions.

*

A similar study by Cushman *et. al.* replicates the paradigm described above by Hauser *et. al.*, again exploring the roles of conscious reasoning and intuition in moral judgment.[65] Pairs of scenarios displaying different variations of the trolley problem were presented to subjects through a test website and, after they had rated the action or omission in each scenario on a 1-to-7 Likert scale from forbidden to obligatory, they were invited to explain the reasons for their decisions.[66] The different scenarios were designed to test for the three typical justifications which were sub-grouped within the first category of sufficient justifications in the study by Hauser *et. al.*, above. These were the 'action principle' whereby harm caused by action is morally worse than omission; the 'intention principle' whereby intended harm is worse than that which is foreseen but unintended; and the 'contact principle' whereby harm caused through physical contact is worse than that inflicted impersonally.[67] However, a greater number of subcategories were employed to distinguish between justifications which were considered sufficient (identifying a relevant factual difference), failed (suggesting an alternative principle but one which does

[63] *Ibid*., 17.
[64] *Ibid*., 17 – 18.
[65] Cushman, Young and Hauser (2006); see also Fiery Cushman and Liane Young, 'Patterns of moral judgment derive from nonmoral psychological representation' (2011) 35 (6) *Cognitive Science: A Multidisciplinary Journal* 1052.
[66] Cushman, Young and Hauser (2006), 1083 – 1084.
[67] *Ibid*., 1083.

not account for the pattern of judgment), uncertain (unable to justify), denying (considering there to be no moral difference between the scenarios despite their prior rating), and alternative explanations introducing unsupported assumptions which were not supported in the scenario.[68]

Cushman *et. al.* found a contrasting range of results. For example, in scenarios testing for the action principle, 'a large majority of subjects were able to provide sufficient justifications for their judgments, whereas relatively few provided failed justifications.' On the one hand, this may indicate in favour of the conscious reasoning model discussed above; on the other, however, the authors equally raise the possibility that subjects constructed the principle *post hoc* upon being asked to justify their responses. Whilst the results cannot distinguish between these positions, they display that a large number of subjects possessed the requisite explicit knowledge that would be required by the conscious-reasoning account.[69] In notable contrast, less than one-third of participants were able to provide sufficient justifications when their pattern of judgments concerned the intention principle, whilst 22% indicated uncertainty in justifying their judgments, 17% denied any morally relevant difference in the scenarios, and 16% 'failed to account for the subject's pattern of judgments.'[70] Responses regarding the contact principle took an intermediate position; interestingly, subjects could typically articulate the relevant principle but were relatively unwilling to endorse it.

The authors consider that, 'although a conscious-reasoning interpretation of subjects' justifications for contact principle cases cannot be rejected definitively, the data favour the intuitionist view.'[71] The above responses concerning the contact principle were particularly informative, as subjects appeared to be using a principle to distinguish between scenarios and yet were unwilling to endorse that same principle when giving reasons for their decisions. As the authors ask, if the conscious reasoning perspective on moral decision-making is correct, 'why would a subject reason consciously from an explicit principle about physical contact during judgment, but then disavow the same

[68] *Ibid.*, 1084 – 1085.
[69] *Ibid.*, 1086.
[70] *Ibid*.
[71] *Ibid.*, 1087.

principle during justification?'[72] Instead, they consider it more likely that the contact principle is guiding decisions following an intuitionist model of moral judgment, after which a 'process of *post hoc* reasoning at justification allows subjects to deduce the principle behind their judgments.'[73] Research concerning the *post hoc* justification for decisions and actions is discussed further in section 7.3.2, below.

Cushman *et. al.* further found that subjects were almost three times as likely to use unsupported alternative explanations with regards to the intention principle as compared with the action principle. They conclude that this demonstrates confabulations which are generated when the subject is asked to justify their decision and finds themselves unable to do so by appeal to the relevant principle being tested. They cite further work by Wheatley and Haidt,[74] which found that subjects' confabulations 'accompanied their inability to provide a principled justification of moral judgment.'[75] Overall, this work confirms many of the findings by Hauser *et. al.*, supporting an intuitionist model of moral decision-making and suggesting a notably reduced involvement of conscious reasoning in the decisions being considered. This work further demonstrates the ready use of confabulation when subjects are otherwise unable to account for their decisions.

### *7.2.1. The Social Intuitionist Model of Moral Judgment*

A large body of research by Jonathan Haidt and various colleagues conducted in 1993,[76] 2000,[77] 2001[78] and beyond, explores moral reasoning and, in particular, posits a "Social Intuitionist Model" of moral decision-making to replace the *status quo* which has been

[72] *Ibid*.
[73] *Ibid*.
[74] Thalia Wheatley and Jonathan Haidt, 'Hypnotic disgust makes moral judgments more severe' (2005) 16(10) *Psychological Science* 780.
[75] Cushman *et. al.* (2006), 1086.
[76] Jonathan Haidt, Silvia Helena Koller and Maria G. Dias, 'Affect, culture, and morality, or is it wrong to eat your dog?' (1993) 65(4) *Journal of Personality and Social Psychology* 613.
[77] Jonathan Haidt, Frederik Björklund and Scott Murphy, 'Moral dumbfounding: When intuition finds no reason' (2000) 1(2) *Lund Psychological Reports* 29.
[78] Jonathan Haidt and Matthew A. Hersh, 'Sexual morality: The cultures and emotions of conservatives and liberals' (2001) 31(1) *Journal of Applied Social Psychology* 191.

dominated by rationalist models.[79] Prior research purported to demonstrate the importance of reasoning and "informational assumptions" in forming moral judgments; for example, people who believe that life begins at conception were more likely to be generally opposed to abortion, whilst those believing that life begins later were generally not opposed.[80] However, Haidt contends that an 'intuitionist approach is just as plausible: the anti-abortion judgment (a gut feeling that abortion is bad) causes the belief that life begins at conception (an *ex post facto* rationalization of the gut feeling).'[81]

Amongst the key research informing Haidt's proposal are a number of experiments where subjects are presented with hypothetical scenarios or actions which are carefully designed to be morally offensive but harmless, such as eating a pet dog, cleaning a toilet with the national flag, or eating a chicken carcass that has been used for masturbation.[82] In what has become one of the most famous examples, the following vignette is presented:

> 'Julie and Mark, who are brother and sister, are traveling together in France. They are both on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy it, but they decide not to do it again. They keep that night as a special secret between them, which makes them feel even closer to each other. So, what do you think about this? Was it wrong for them to have sex?'[83]

Data was drawn from Likert scale self-reports, behaviour videotaped during each task, and demographic and personal information provided by each subject. Participants

[79] See further Jonathan Haidt, 'The emotional dog and its rational tail: A social intuitionist approach to moral judgment' (2001) 108(4) *Psychological Review* 814; Jonathan Haidt, 'The new synthesis in moral psychology' (2007) 316(5827) *Science* 998.
[80] Elliot Turiel, Carolyn Hildebrandt, Cecilia Wainryb and Herbert D. Saltzstein, 'Judging social issues: Difficulties, inconsistencies, and consistencies' (1991) 56(2) *Monographs of the Society for Research in Child Development* 1.
[81] Haidt (2001), 817.
[82] Haidt, Koller and Dias (1993).
[83] Haidt, Björklund and Murphy (2000), 18.

generally considered the actions in each scenario to be 'wrong, and universally wrong',[84] and frequently offered statements such as '"It's just wrong to do that!" or "That's terrible!".'[85] It was further found that affective reactions (expressions or statements of distaste or disgust) were better predictors of subjects' moral predictions than their claims regarding potentially harmful consequences from each scenario.[86] Moreover, subjects were often "morally dumbfounded"; 'that is, they would stutter, laugh, and express surprise at their inability to find supporting reasons, yet they would not change their initial judgments of condemnation.'[87]

Haidt posits that the 'central claim of the social intuitionist model is that moral judgment is caused by quick moral intuitions, and is subsequently followed (when needed) by slow, *ex post facto* moral reasoning.'[88] Furthermore, Haidt proposes that moral reasoning is a fundamentally social exercise for explaining what somebody has done and why they have done it, within a social context. From this perspective, moral reasoning does not operate for the purpose of discovering moral truths but, rather, acts like a 'lawyer or politician seeking whatever is useful, whether or not it is true,' to explain an individual's actions within a social setting.[89] This would explain why people generally have very rapid responses to moral violations, why affective reactions are better predictors of moral judgments, and why people can often become morally dumbfounded when they are unable to reconcile their moral intuitions with concrete reasons against a particular action or scenario. More broadly, Haidt cites studies of 'everyday reasoning' which similarly suggest that people generally approach reasoning by 'setting out to confirm their initial hypothesis.'[90]

[84] Haidt (2001), 817.
[85] Haidt, Björklund and Murphy (2000), 9.
[86] Haidt (2001), 817.
[87] *Ibid*.
[88] *Ibid*., 817.
[89] Haidt (2007), 999.
[90] *Ibid*., 998; citing Deanna Kuhn, *The Skills of Argument* (Cambridge University Press 1991).

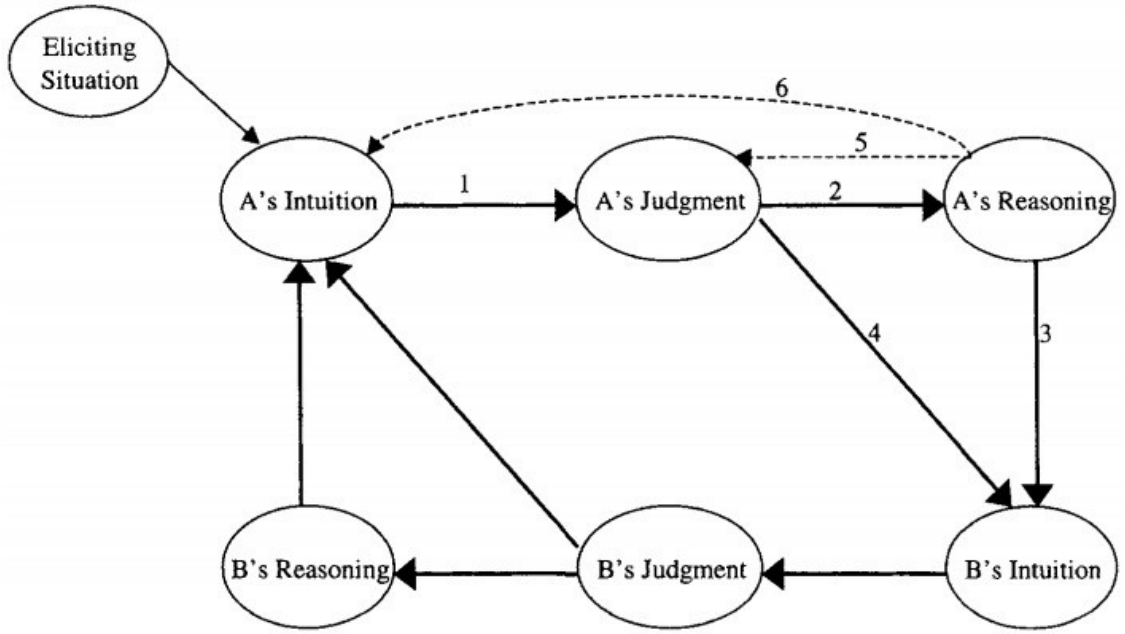


***Fig. k – Social intuitionist model of moral judgment; (1) intuitive judgment link, (2)* post hoc *reasoning link, (3) reasoned persuasion link, (4) social persuasion link, (5) reasoned judgment link, (6) private reflection link.*[91]**

The various numbered links in ***figure k***, above, represent a number of propositions under the social intuitionist model. The first link (1) proposes that 'moral judgments appear in consciousness automatically and effortlessly as the result of moral *intuitions*.'[92] This concurs with other examples suggesting that a large proportion of social cognition 'operates automatically and implicitly',[93] many of which are considered throughout this thesis such as the range of experiments by John Bargh and colleagues discussed in section 3.1, above. The second link (2) proposes that moral reasoning is an effortful process taking place after a moral judgment has first been made, 'in which a person searches for arguments that will support an already-made judgment', drawing from evidence that everyday reasoning is biased by the search for supporting evidence.[94] The third link (3) proposes that moral reasoning is 'produced and sent forth verbally to justify one's already-made moral judgment to others,' introducing an inherently social element to the

[91] *Ibid*., 815.
[92] Haidt (2001), 818 (emphasis added).
[93] *Ibid*; citing John A. Bargh and Tanya L. Chartrand, 'The unbearable automaticity of being' (1999) 54(7) *American Psychologist* 462; Anthony G. Greenwald and Mahzarin R. Banaji, 'Implicit social cognition: Attitudes, self-esteem, and stereotypes' (1995) 102(1) *Psychological Review* 4; see also Anthony G. Greenwald and Calvin K. Lai, 'Implicit social cognition' (2020) 71(25) *Annual Review of Psychology* 1.
[94] Haidt (2001), 818; citing Nisbett Wilson (1977); Kuhn (1991); Ziva Kunda, 'The case for motivated reasoning' (1990) 108(3) *Psychological Bulletin* 480; David N. Perkins, Michael Faraday and Barbara Bushey, 'Everyday reasoning and the rots of intelligence' in Voss J. F., Perkins D. N and Segal J. W. (eds.), *Informal Reasoning and Education* (Lawrence Erlbaum Associates 1991).

model.[95] Haidt suggests in particular that although moral reasoning can sometimes impact upon other people's judgments, generally moral arguments are intractable and 'notorious for the rarity with which persuasion takes place.'[96] Just as affective reactions are a more robust predictor of moral attitudes, so evidence highlights the importance of deploying affective persuasion to effectively change those attitudes, further suggesting that moral judgments are generally more intuitive.[97]

The fourth link (4) proposes that the moral judgments of friends, family, colleagues and acquaintances (*i.e.,* one's social circle and society in general) exert a direct influence on the judgments of others without the need for reasoning or persuasion to be used, simply because people are 'highly attuned to the emergence of group norms.'[98] This introduces a further inherently social element to the model, suggesting that social forces may not only elicit outward conformity from others but also directly shape their privately held judgments. Haidt explains that these first four links form the "core" of the social intuitionist model, giving moral reasoning a 'causal role in moral judgment only when reasoning runs through other people.'[99] The final two links hypothesise the means through which private reflection can impact upon moral judgments; however, it is suggested that these 'rarely override [people's] initial intuitive judgments just by reasoning privately to themselves because reasoning is rarely used to question one's own attitudes or beliefs.'[100]

The fifth link (5) proposes that people may sometimes be able to override their initial intuition through sheer reason and logic, but that such a capacity would typically be rare and occurring primarily when a given moral intuition was weak and mental processing capacity high. Haidt cites the work of Wilson, Lindsay and Schooler[101] which suggests

---

[95] Haidt (2001), 818 – 819.
[96] *Ibid*., 819.
[97] *Ibid*; citing Kari Edwards and William von Hippel, 'Hearts and minds: The priority of affective versus cognitive factors in person perception' (1995) 21(10) *Personality and Social Psychology Bulletin* 996; Sharon Shavitt, 'The role of attitude objects in attitude functions' (1990) 26(2) *Journal of Experimental Social Psychology* 124.
[98] Haidt (2001), 819.
[99] *Ibid*.
[100] *Ibid*.
[101] Timothy D. Wilson, Samuel Lindsey and Tonya Y. Schooler, 'A modal of dual attitudes' (2000) 107(1) *Psychological Review* 101.

that the conflict between a strong intuition and reasoned judgment may result in a "dual attitude" in which the 'reasoned judgment may be expressed verbally yet the intuitive judgment continues to exist under the surface.'[102] It might further be suggested that the dual attitude similarly describes the phenomenon of being morally dumbfounded, discussed above. Applying this reasoning, the morally dumbfounded individual experiences a strong intuitive judgment against a particular action or scenario, despite being unable to adequately explain or justify this judgment through logic or reasoning.

The sixth and final link (6) proposes that the process of deliberating about a particular situation may 'spontaneously activate a new intuition that contradicts the initial intuitive judgment.'[103] In essence, this is the process by which somebody "puts themselves in another's shoes" and, in so doing, may appreciate a particular dilemma or circumstance from a different perspective and thereby experience conflicting intuitions about how the judge that situation. Crucially, Haidt notes, rationalist theories of moral decision-making have emphasised the latter two links whereas, in contrast, intuitionist theories emphasise the first four links, whilst allowing the final two to offer some contribution on rarer occasions.[104]

The social intuitionist model of moral decision-making posited by Haidt may be reflected in role of consciousness in the process of deliberating a decision over time that has been hypothesised throughout this thesis. Referring to ***figure d*** of section 2.3.1 of this thesis, an intuitive judgment made rapidly and without conscious consideration might be reflected by the horizontal line to the left of the figure at which point option B is selected. This represents a decision that is taken rapidly and intuitively, the winning option being that which most rapidly attracts the greatest valence amongst the competing neuronal bundles representing each decision option. This decision will likely only engage the first four links described from the social intuitionist model.

[102] Haidt (2001), 819.
[103] *Ibid*.
[104] *Ibid*.

However, the conscious deliberation of the same decision offers greater time and mental resources to that decision-making process, which occurs when an individual is required to provide reasons for their decision and, crucially, is likely necessary in order to engage links five and six of the social intuitionist model. At this stage, moral reasoning *may* override initial intuitions because, referring again to ***figure d***, more time and greater resources have been made available in order to gather evidence in favour of option A instead of option B. Critically, it is hypothesised that the competing networks of neurons representing different decision options which provide the source of the initial moral intuition, are the same competing networks which provide the source of a moral judgment produced by reasoning over time. The only real difference between the two scenarios, it is posited, are the increased time and mental resources afforded to those mutually competing networks by the process of conscious deliberation.

The greater significance of the social intuitionist model of moral judgments to the *why* component of decision-making and legal responsibility generally, again, lies in the apparent disconnection between decisions that are taken on the one hand, and reasons that are given in justification or explanation for those decisions on the other. If, following a rationalist model, moral decisions were reached on the basis of reasoning from first principles and applying them to a given set of circumstances, the expectation would be that people could reliably and accurately recall the underlying principles or reasons that have informed and resulted in their decisions. Conversely, the evidence explored in this section supports an intuitionist model, strongly suggesting that moral reasoning occurs *post hoc* as an exercise for rationalising and explaining moral decisions, which are first reach intuitively.

This further explains why there appears to be such a disconnection between genuine reasons for decisions and the reasons that can typically be subjectively accessed by individuals, discussed in sections 7.1.1 to 7.1.3 of this chapter. The implications of this intuitionist model for legal responsibility, however, are that individuals appear to have poor subjective access to the genuine reasons motivating their decisions and behaviour, and can readily confabulate reasons to fill this explanatory gap, concurrent with previous discussions in this chapter. Consequently, the courtroom inquiry into why somebody

performed a particular action, which is invariably deployed by both prosecution and defence in order to elicit and negate *mens rea* respectively, can only yield unreliable answers, rendering the current approach to *mens rea* an unreliable means of determining responsibility.

### *7.2.2. Universal Moral Grammar*

Although offering a compelling account of how people reach rapid and automatic moral judgments whilst remaining deficient in explaining the underlying reasons thereof, social intuitionist models of moral decision-making leave certain queries unanswered. Such questions include how an intuitionist decision-making system would yield relatively consistent patterns of answers to scenarios such as the trolley problem across vastly diverse cultures and societies around the world, or how reason and argument might interact with an intuitionist system in order to change people's beliefs and opinions about certain moral and / or legal problems. One potential solution to these and similar questions is provided by the theory of "Universal Moral Grammar" ('UMG') elaborated by John Mikhail.[105]

Drawing inspiration from Noam Chomsky's theory of universal grammar in linguistics, UMG proposes that deontic moral knowledge consists of mental structures containing 'a system of rules and principles that generates and relates mental representations of various types… [and] is what enables individuals to distinguish actions that are morally permissible from those that are not.'[106] In this regard, permissibility judgments depend not upon superficial properties of a particular problem or scenario but upon the way it is represented in the brain. Further, UMG proposes that 'at least some operative moral principles are inaccessible to consciousness suggest[ing] that, as in the case with language,

[105] John M. Mikhail, 'Rawls' linguistic analogy: A study of the "generative grammar" model of moral theory described by John Rawls in *A Theory of Justice*' (DPhil thesis, Cornell University 2000); John M. Mikhail, *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment* (Cambridge University Press 2011).

[106] John M. Mikhail, Cristina M. Sorrentino and Elizabeth S. Spelke, 'Toward a universal moral grammar' in Gernsbacher M. A. and Derry S. J. (eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (Lawrence Erlbaum Associates 1998), 1250.

these principles are not taught to successive generations explicitly… [but] are the developmental consequences of an innate, cognitive faculty.'[107]

Initial evidence in favour of UMG is drawn from a variety of observations and experiments.[108] For example, young children display relatively complex moral intuitions that are not readily explained by teaching and learned experience alone. Children aged three to four years can distinguish between two acts which have the same consequences according to their different purposes or intentions,[109] and can also distinguish genuine moral violations (such as battery and theft) from breaches of social conventions (such as wearing pyjamas to school).[110] At four- to five-years-old, children apply the principle of proportionality to attribute the correct relative degrees of punishment between principals and accessories,[111] and five- to six-year-old children will exculpate behaviours based on incorrect factual belief but not false moral beliefs.[112] Such findings support the propositions of UMG given the relative paucity of exposure to either complex moral problems or explicit training in moral principles at these ages.

Furthermore, it is observed with interest that all natural languages in the world include words or phrases expressing the same basic moral concepts such as obligation, forbiddance and permissibility.[113] In a similar vein, the vast majority of legal systems around the world appear to proscribe against a number of similar offences – and violent / aggressive offences in particular – such as murder, rape and battery,[114] and rely on a number of similar distinctions in assessing responsibility for actions such as causation,

---

[107] *Ibid.*

[108] See John M. Mikhail, 'Universal moral grammar: Theory, evidence and the future' (2007) 11(4) *Trends in Cognitive Sciences* 143, 143 – 144.

[109] Sharon A. Nelson, 'Factors influencing young children's use of motives and outcomes as moral criteria' (1980) 51(3) *Child Development* 823.

[110] Judith G. Smetana, 'Social-cognitive development: Domain distinctions and coordinations' (1983) 3(2) *Developmental Review* 131.

[111] Norman J. Finkel, Marsha B. Liss and Virginia R. Moran, 'Equal of proportional justice for accessories? Children's pearls of proportionate wisdom' (1997) 18(2) *Journal of Applied Developmental Psychology* 229.

[112] Michael J. Chandler, Bryan W. Sokol and Cecilia Wainryb, 'Beliefs about truth and beliefs about rightness' (2000) 71(1) *Child Development* 91.

[113] Joan L. Bybee and Suzanne Fleischman (eds.), *Modality in Grammar and Discourse* (John Benjamins Publishing 1995).

[114] Donald Brown, *Human Universals* (McGraw-Hill Companies 1991); John M. Mikhail, 'Law, science, and morality: A review of Richard Posner's "The problematics of moral and legal theory"' (2002) 54(5) *Stanford Law Review* 1057.

intentionality and voluntariness.[115] In this regard, Mikhail argues that scholars of comparative law have 'suggested that a few basic distinctions capture the "universal grammar" of all systems of criminal law.'[116] Finally, the observation of common brain circuitry that appears to be engaged in the process of moral reasoning is further read as suggesting towards a similarly common (or "universal") moral grammar which that brain circuitry applies in all (neurotypical) people.

In addition to providing some explanation for the evidential findings considered above, UMG is argued to support two further fundamental arguments.[117] First, the "argument for moral grammar" expresses the suggestion that the mind contains a moral grammar analogous to the linguistic grammar that enables people to learn languages, recognise well-formed sentences, and go on to create entirely novel phrases through the application of a limited number of grammatical rules. Thus, the brain's moral grammar consists of a 'complex and possibly domain-specific set of rules, concepts and principles that generates and relates mental representations of various types… [and] enables individuals to determine the deontic status of an infinite variety of acts and omissions.'[118] Second, the "argument from the poverty of the moral stimulus" reflects on the ability for people to form moral judgments regarding a theoretically infinite range of novel scenarios, despite the comparatively limited number of such scenarios that are actually encountered during life, and not least during early development when children are learning to apply different moral concepts. The suggestion follows that:

> '[T]he manner in which this grammar is acquired implies that at least some of its core attributes are innate, where "innate" is used in a dispositional sense to refer to cognitive systems whose essential properties are largely pre-determined by the inherent structure of the mind, but whose ontogenetic development must be triggered and shaped by appropriate

[115] Mikhail (2002); George P. Fletcher, *Basic Concepts of Criminal Law* (Oxford University Press 1998); Stuart P. Green, 'The universal grammar of criminal law' (2000) 98(6) *Michigan Law Review* 2104.
[116] Mikhail (2007), 143; citing Fletcher (1998); Green (2000).
[117] *Ibid.,* 144.
[118] *Ibid*; citing Mikhail (2000).

> experience and can be impeded by unusually hostile learning environments.'[119]

Mikhail proceeds to offer a number of principles that are likely candidates for existing within the innate moral grammar structures of the brain.[120] For example, the "principle of natural liberty" suggests that all types of actions are *prima facie* permitted unless explicitly defined and forbidden by some rule, which underpins the maxim of there being no crime / punishment without law contained within the legal concept of the rule of law.[121] The "prohibition of battery and homicide" eponymously states that unpermitted and unprivileged contact with another is forbidden, ranging from mere touching (battery) to killing other people.[122] The "principle of self-preservation" suggests that it is generally permissible to attempt to protect oneself from harm or death and, broadly speaking, otherwise prohibited battery against another may be permissible to protect that other from more harmful consequences, applying the (rebuttable) presumption that the other would similarly consent to be protected from some greater harm.[123]

At this juncture, the work of Patricia Churchland [124] is potentially relevant for understanding from where such innate mental structures representing moral principles in the brain might arise. In very brief terms, Churchland argues that moral intuitions have profoundly (neuro)biological origins and, more broadly, the experience of moral conscience results from the manner in which the human (or, more generally, primate) brain has evolved within a social context. Morality, therefore, is a 'natural phenomenon

---

[119] *Ibid*; citing Mikhail (1998); Mikhail (2000); Noam Chomsky, *Knowledge of Language: Its Nature, Origin, and Use* (Praeger 1986); Noam Chomsky, *The Minimalist Program* (Massachusetts Institutes of Technology Press 1995); Ray Jackendoff, *Patterns in the Mind: Language and Human Nature* (Basic Books 1995); Charles R. Gallistel, 'The replacement of general-purpose learning models with adaptively specialized learning modules' in Gazzaniga M. S. (ed.), *The Cognitive Neurosciences* (2nd ed. Massachusetts Institute of Technology Press 2000).

[120] John M. Mikhail, 'Moral grammar and intuitive jurisprudence: A formal model of unconscious moral and legal knowledge' in Bartels D. M., Bauman C. W., Skitka L. J. and Medin D. L. (eds.), *Psychology of Learning and Motivation: Moral Judgment and Decision Making* (Academic Press 2009).

[121] *Ibid.,* 52.

[122] *Ibid*., 53 – 55.

[123] *Ibid*., 55 – 56.

[124] Patricia S. Churchland, *Braintrust: What Neuroscience Tells Us about Morality* (Princeton University Press 2011); Patricia S. Churchland, *Conscience: The Origins of Moral Intuition* (W. W. Norton & Company 2019).

– constrained by the forces of natural selection, rooted in neurobiology, shaped by the local ecology, and modified by cultural developments.' [125] Oxytocin is a hormone significantly responsible for mammalian bonding and is attributed with fostering care, attachment and trust between creatures, principally mothers and children. Within social creatures such as the great apes, the effects of oxytocin extend beyond immediate parent-infant relations and similarly contribute to bonding between less closely related creatures living in the same broader social family or group. Further, the intricate neural circuitry for pain and reward extend to contribute to the pain of separation from, and the enjoyment of company with, social animals within one's own group, not least for great apes and humans; 'the pain of exclusion, separation, and disapproval, … exploits, expands, and modifies what is already in place for physical pain and homeostatic emotions in premammalian species.' [126] It is from such neurobiological process, evolving and operating within a social context, that Churchland proposes innate moral values emerge, arising from attachment to family, caring for more distant friends, and the need to belong within a wider social group: 'attachment begets caring; caring begets conscience.'[127]

It is not difficult to hypothesise how such (neuro)biological processes may result in many of the "innate" moral principles postulated within Mikhail's universal moral grammar. For example, the unpleasantness of pain (mediated *inter alia* by the activation of pain neurons and the hormonal hypothalamic-pituitary-adrenal-thyroid-gonadal axis) and the enjoyment of pleasure (mediated *inter alia* by the dopaminergic reward system) generally causes all animals to avoid the former and act in pursuit of the latter. In social animals, however, social rules often constrain individuals from being able to act unreservedly in this way; for example, social hierarchies will often restrict the order in which individuals can feed and how much food they receive, whether and how frequently they are able to mate, and whether and how frequently they are groomed by others, *etc.* If this pursuit of pleasure within the constraints of social hierarchies were to be formalised into a deontic rule and expressed by dedicated brain circuitry, it might appear something like the

[125] Churchland (2011), 191.
[126] *Ibid.,* 46.
[127] Churchland (2019), 49.

aforementioned principle of natural liberty whereunder an individual can act in any way it wishes save for those actions that are forbidden by the rules of its social environment.

Similarly, unwanted touching will arouse stress and anxiety (mediated by the hypothalamic-pituitary-adrenal axis and the hormone cortisol) in most animals, whilst more violent touching will cause aversive pain. In a social context where peaceful and mutually beneficial interactions are optimal for individuals, the formalisation into a deontic rule of this mutual aversion to stress and pain might appear something like the aforementioned principle of the prohibition of battery and homicide. Extending the argument further still, the typical biological response to unwanted touching and violent attack in particular is the "fight or flight" response (mediated by activation of the amygdala and the hormone adrenaline), which exists principally for the purposes of self-preservation. Again, formalising such a biological response into a deontic rule for application within a social setting may result in something akin to the principle of self-preservation. Thus, it is hypothesised that the evolution of such biological mechanisms within a social context as discussed by Churchland, may provide the genesis for the innate brain structures and associated moral principles developed within Mikhail's universal moral grammar.

Returning to the UMG theory, a number of final points must be made with particular relevance to the present thesis. First, in concurrence with the social intuitionist model of moral decision-making and, indeed, the majority of mental processes considered in this thesis, UMG proposes that moral intuitions are arrived at automatically; 'they are not made by a conscious application of moral rules or principles.'[128] This does not mean that moral judgments are themselves unprincipled; rather, the proposition is that the brain possesses 'tacit or unconscious knowledge of a rich variety of legal rules, concepts, and principles, along with a natural readiness to compute mental representations of human acts and omissions in legally cognizable terms.'[129] Here, again, the analogy to linguistic grammar is drawn, whereby the brain appears to intuitively grasp grammar during the process of language learning, and then automatically applies that grammar to both

[128] Mikhail (2011), 82.
[129] Mikhail (2009), 29.

recognise and create grammatically correct sentences. Nonetheless, despite the automaticity of the process, the subsequent production of correct language follows a tacitly understood collection of rules and principles.

Second, and relatedly, the theory of UMG:

> '[D]istinguish[es] sharply between an individual's *operative* moral principles (those principles actually operative in her exercise of moral judgment) and her *express* principles (those statements she makes in the attempt to describe, explain, or justify her judgments). We make no assumption that the normal individual is aware of the operative principles which constitute her moral knowledge, or that she can become aware of them through introspection, or that her statements about them are necessarily accurate. On the contrary, we hypothesize that just as normal persons are typically unaware of the principles guiding their linguistic or visual intuitions, so too are they often unaware of the principles guiding their moral intuitions.'[130]

This proposition flows from the body of research considered throughout this chapter and in section 7.2 in particular, above, demonstrating the relatively poor access to genuine reasons for decisions that people appear to have and, in the absence of such access, the ready propensity for the brain to confabulate such reasons.

Third, it is recalled that UMG proposes that the brain possesses innate structures which enable it to ontogenetically learn moral grammar from a relative paucity of experience with moral problems and teaching in early development. In other words, just as structures in the brain enable children to automatically learn the language(s) to which they are exposed during early development, so it is hypothesised that analogous structures enable the automatic learning of moral grammar to which people are similarly exposed. The

[130] John M. Mikhail, 'Aspects of the theory of moral cognition: Investigating intuitive knowledge of the prohibition of intentional battery and the principle of double effect' (Georgetown University Law Center, Working paper no. 762385, 2002), 3 – 4 (original emphasis).

crucial point in this regard is twofold; on the one hand, the developing brain must be "exposed" to moral principles in the first place, whether through the passive observation of events and scenarios, active participation therein, or through explicit learning taught by parents and other adults. On the other hand, such education and learning does not simply cease upon reaching maturity; just as adults can learn new languages (albeit with greater difficulty than children), so too the adult brain continues to learn, develop and refine its moral grammar.

*

A legitimate question asks whether, how, and to what extent can people use and apply reason in a decision-making process that appears to be automatic and intuitive; put differently, how do people respond to reason in moral decision-making if the processes involved are unconscious and automatic? A number of responses are forthcoming. In the first instance, and following from the immediately preceding paragraph, the initial moral education of a child's brain will virtually always include any number of perfectly rational principles – caring for and being helpful to others within one's family, school class or other social circle; not hitting others (prohibition of battery); seeking an adult if in danger or under attack (principle of self-preservation), *etc.* As the proficiency with which a person applies such principles to novel situations increases (and, perhaps crucially, is rewarded by their social environment), so the application of such principles becomes increasingly automatic. This argument follows a more general conception in neuropsychology stating that 'complex cognitive operations eventually migrate from System 2 to System 1 as proficiency and skill are acquired,'[131] where "system 2" refers to effortful, conscious and controlled actions such as learning to drive or play a music instrument, and "system 1" refers to effortless, unconscious and automatic processes such as when a *learned and practised* driver or instrumentalist performs that particular skill.

[131] Daniel Kahneman and Shane Frederick, 'Representativeness revisited: Attribute substitution in intuitive judgment' in Gilovich T., Griffin D. and Kahneman D. (eds.), *Heuristics and Biases* (Cambridge University Press 2002), 51.

Lovibond reflects on the social activities in which humans participate, including encounters with activities, scenarios and circumstances of a moral character:

> 'Over time, our participation in these activities – while creating a succession of new contexts for thought and decision – gives rise to a "second," or acquired, nature. This second nature is manifested in behaviour which, though learned, is largely unreflective (like the speaking of a first language); and which, if we do make it into an object of reflection, usually produces in us a sense of inevitability. From one point of view, the dispositions that constitute our second nature are passive, for they are dispositions to be affected in a certain way: ideally, to register the "proper force and necessity" of reasons for judgment (or for action). However, a feature of human socialization – of the sum of "activities" in the simple… sense into which we (humans) have been initiated – that one is led not just to receive and process sensory input from one's environment, but to recognize the state of the world as imposing rational constraints on one's thinking. And the dawning of this recognition is what… enlists us as participants in the "*active* adjustment" of thought to world.'[132]

Where UMG proposes that the brain contains innate structures which enable it to recognise and learn moral grammar from its social and cultural environment, analogous to the learning of language, exposure to that environment is therefore crucial to any brain's moral education. Furthermore, as any individual learns rational principles through passive observation, active experience or explicit teaching, and practices the application of those principles (encouraged by social reinforcement), so they become increasingly automatic and intuitive; 'in the process of moral upbringing, rational grounds become embodied in our intuitive thinking.'[133]

---

[132] Sabina Lovibond, *Ethical Formation* (Harvard University Press 2002), 25 – 26 (emphasis added); citing John Henry McDowell, *Mind and World* (Harvard University Press 1994), 84.

[133] Hanno Sauer, 'Education institution, automaticity and rationality in moral judgment' (2012) 15(3) *Philosophical Explorations* 255, 256.

This argument only takes the discussion part of the way, however, for even if intuitive moral judgment applies previously learned and perfectly rational rules and principles, the question remains how these automatic systems might be updated in adulthood, for example, when a particular (automatically operative) moral principle is not applicable to a given novel situation, or when people update and change their beliefs and values throughout their life. Fortunately in this regard, learning and education does not cease upon maturity of the brain, and people are clearly capable of learning, adopting and applying new moral lessons and principles throughout their life (just as people can learn a second language in adulthood). For example, Sauer argues that whilst the actualisation and execution of intuitive moral judgments may take place on a subconscious level, this does not mean that an agent's practical reasons for actions cannot also operate on a subconscious level.[134] Thus, when an individual encounters new, improved or otherwise persuasive (*i.e.,* "good") reasons to do or not do a particular thing (or make a particular judgment), the automaticity of decision-making processes does not preclude the inclusion of these new reasons, but only suggests that such inclusion may itself occur automatically and beneath the level of conscious awareness.

Taking the argument that learning and education continues throughout maturity and can thus impact upon and change underlying automatic processes, at least two broad categories of learning might be adopted to influence and change otherwise automatic moral intuitions – *ex ante* education and *ex post* education. The former *ex ante* education is concerned with antecedents to the generation of a moral intuition:

> 'Prior reasoning can determine the sorts of output that emerge from intuitive systems. This can happen through shifts in cognitive appraisal, as well as through conscious decisions as to what situations to expose oneself to. In both of these regards, prior controlled processes partially determine which fast, unconscious, and automatic intuitions emerge.'[135]

[134] *Ibid*., 263.
[135] David A. Pizarro and Paul Bloom, 'The intelligence of the moral intuition: Comment on Haidt' (2003) 110(1) *Psychological Review* 193, 194.

In a negative sense, people can avoid the generation of unwanted moral intuitions by avoiding situations in which they might expect those intuitions to be generated. In a more positive sense, however, people can selectively expose themselves to situational stimuli in order to develop or reinforce desired moral intuitions; for example, somebody wanting to adopt a "more moral" approach to food and diet by becoming vegetarian or vegan may start reading and watching materials about the horrors of the meat industry and factory farming.[136] In this regard, research shows that deliberately and positively interacting with people of different races (as well as other prejudiced groups such homosexuals,[137] people with disabilities,[138] and people with mental illness[139]) can be an effective strategy for counteracting unwanted prejudicial attitudes and biases that intuitively arise in relation to that group of people.[140]

One prominent example of an *ex ante* implementation strategy consists of *if-then* plans, whereby individuals decide in advice to respond in a particular way if and when they

[136] Sauer (2012), 267; see also Jeanette Kennett and Cordelia Fine, 'Will the real moral judgment please stand up? The implications of social intuitionist models of cognition for meta-ethics and moral psychology' (2009) 12(1) *Ethical Theory and Moral Practice* 77, 91 – 93.

[137] Amy B. Becker, 'Determinants of public support for same-sex marriage: Generational cohorts, social contact, and shifting attitudes' (2012) 24(4) *International Journal of Public Opinion Research* 524; Gregory M. Herek and John P. Capitanio, '"Some of my best friends": Intergroup contact, concealable stigma, and heterosexuals' attitudes toward gay men and lesbians' (1996) 22(4) *Personality and Social Psychology Bulletin* 412.

[138] Lynn Anderson, Stuart J. Schleien, Leo McAvoy, Greg Lais and Deborah Seligmann, 'Creating positive change through an integrated outdoor adventure program' (1997) 31(4) *Therapeutic Recreation Journal* 214.

[139] Laurel Alexander and Bruce Link, 'The impact of contact on stigmatizing attitudes toward people with mental illness' (2003) 12(3) *Journal of Mental Health* 271; Donna M. Desforges, Charles G. Lord, Sherri L. Ramsey, Jonathan A. Mason, M. D. van Leeuwen, Stephen C. West and Mark R. Leper, 'Effects of structured cooperative contact on changing negative attitudes toward stigmatized social groups' (1991) 60(4) *Journal of Personality and Social Psychology* 531.

[140] Thomas F. Pettigrew and Linda R. Tropp, 'A meta-analytic test of intergroup contact theory' (2006) 90(5) *Journal of Personality and Social Psychology* 751; Thomas F. Pettigrew, 'Intergroup contact theory' (1998) 49(1) *Annual Review of Psychology* 65; Elirea Bornman and Johan C. Mynhardt, 'Social identity and intergroup contact with specific reference to the work situation' (1991) 117(4) *Genetic, Social, and General Psychology Monographs* 437; Hwa-Bao Chang, 'Attitudes of Chinese students in the United States' (1973) 58(1) *Sociology and Social Research* 66; Ernest Works, 'The prejudice-interaction hypothesis from the point of view of the negro minority group' (1961) 67(1) *American Journal of Sociology* 47; Gordon W. Allport, *The Nature of Prejudice* (Addison-Wesley 1954); Daniel M. Wilner, Rosabelle Price Walkley and Stuart W. Cook, *Human Relations in Interracial Housing: A Study of the Contact Hypothesis* (University of Minnesota Press 1955); Morton Deutsch and Mary Evans Collins, *Interracial Housing: A Psychological Evaluation of a Social Experiment* (University of Minnesota Press 1951).

encounter some anticipated stimuli or intuitive response thereto.[141] Gallo *et. al.* found such a strategy to be significantly effective in reducing automatic and highly intuitive emotional responses of disgust and fear to various stimuli in a study where subjects formed *if-then* plans to remain calm and relaxed when they were presented with the disgust- or fear-inducing stimuli.[142] Similarly, in a study deploying a weapon-identification task which measures implicit bias towards the faces of Black men, subjects made significantly fewer false-positive gun identifications in response to the presentation of images of a Black face when they made the commitment, 'whenever I see a Black face on the screen, I will think the word "safe".'[143] Whilst these examples consist of situations where subjects have explicitly been provided with the relevant strategy for moderating or altering their intuitive judgments and responses, further research shows that people with higher capacities for self-regulation and a stronger motivation to control their own prejudices (or other intuitive judgments and responses) 'show less behavioural expression of automatically activated associations', demonstrating how *ex ante* education can occur spontaneously as well as being prompted by others.[144] In this regard, Barrett *et. al.* speculate that:

> '[C]ontrolled processing may not be merely reversing the effects of automatic processing, but it may also prevent (or allow) the expression of attention on representations that were activated in a stimulus-driven way. As long as one has a processing goal (like an egalitarian goal to prevent stereotyping, for example), as well as the [working memory capacity ('WMC')] to deploy goal-directed attentional effects, the processing goal can be enacted. As a result, some of the effects that we think of as

[141] Peter M. Gollwitzer, 'Implementation intentions: Strong effects of simple plans' (1999) 54(7) *American Psychologist* 493; Inge Schweiger Gallo, Andreas Keil, Kathleen C. McCulloch, Brigitte Rockstroh and Peter M. Gollwitzer, 'Strategic automation of emotion regulation' (2009) 96(1) *Journal of Personality and Social Psychology* 11.
[142] Gallo (2009).
[143] Brandon D. Stewart and B. Keith Payne, 'Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control' (2008) 34(10) *Personality and Social Psychology Bulletin* 1332, 1336.
[144] Kennett and Fine (2009) 91; citing B. Keith Payne, 'Conceptualizing control in social cognition: How executive functioning modulates the expression of automatic stereotyping' (2005) 89(4) *Journal of Personality and Social Psychology* 488; David M. Amodio, Patricia G. Devine and Eddie Harmon-Jones, 'Individual differences in the regulation of intergroup bias: The role of conflict monitoring and neural signals for control' (2008) 94(1) *Journal of Personality and Social Psychology* 60.

> automatic… may well involve the control of attention so early on that there is no associated experience of will or agency. For example, it may be that a property of the person (*e.g.,* skin pigmentation) automatically activates both a stereotype and a goal to be egalitarian, and with sufficient WMC resources, the activation level of the stereotype can be suppressed before it influences subsequent processing, thereby allowing egalitarian outcomes with perceived ease.'[145]

The latter *ex post* education is principally concerned with the capacity for metacognition;[146] that is, the ability for people to monitor, reflect on, and (sometimes) alter their own cognitive functions and subsequent outputs – *i.e.,* "thinking about thinking." For example, whilst it is well known that an individual's incidental affective states can impact upon or "contaminate" their subsequent moral judgments,[147] people are capable of correcting for the impact of such transient moods when their attention is drawn to their bias or when they are particularly motivated towards accuracy.[148] Again, such learning need not necessarily be prompted by others, as people are also capable of spontaneously recognising errors, prejudices and biases in their own intuitive judgments and effortfully exercising a degree of control to correct for these errors in subsequent decisions.[149]

Crucially, it is submitted that *ex post* education – and, in particular, spontaneous *ex post* reasoning about moral judgments and decisions – operates under a number of constraints. It requires that an individual recognises some flaw, error, prejudice or bias within their

---

[145] Lisa Feldman Barrett, Michele M. Tugade and Randall W. Engle, 'Individual differences in working memory capacity and dual-process theories of the mind' (2004) 130(4) *Psychological Bulletin* 553, 564.
[146] Sauer (2012), 268.
[147] Joseph P. Forgas and Stephanie Moylan, 'After the movies: Transient mood and social judgments' (1987) 13(4) *Personality and Social Psychology Bulletin* 467; Wheatley and Haidt (2005).
[148] Jennifer S. Lerner, Julie H. Goldberg and Philip E. Tetlock, 'Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility' (1998) 24(6) *Personality and Social Psychology Bulletin* 563; Timothy D. Wilson and Nancy Brekke, 'Mental contamination and mental correction: Unwanted influences on judgments and evaluations' (1994) 116(1) *Psychological Bulletin* 117; Norbert Schwarz and Gerald L. Clore, 'Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states' (1983) 45(3) *Journal of Personality and Social Psychology* 513.
[149] Cordelia Fine, 'Is the emotional dog wagging its rational tail, or chasing it?' (2006) 9(1) *Philosophical Explorations* 83; Margo J. Monteith, Leslie Ashburn-Nardo, Corine I. Voils and Alexander M. Czopp, 'Putting the brakes on prejudice: On the development and operation of cues for control' (2002) 83(5) *Journal of Personality and Social Psychology* 1029; Haidt and Hersh (2001).

moral intuitions in the first place; that the individual is sufficiently motivated to address such an identified error; that they possess the requisite capacities (*i.e.,* executive functions) in order to effect the necessary mental changes; and that they possess the requisite resources (*i.e.,* energy) in order to operate those capacities effectively.[150] All this renders conscious moral reasoning as being comparatively slow and effortful, and an altogether rarer occurrence in relation to many people and most judgments and decisions. Moreover, Sauer argues that moral reasoning does not operate to precede and then cause moral judgments *per se*, but provides feedback into the mechanisms responsible for producing the automatic and intuitive moral judgments. Thus, conscious moral reasoning is:

> '[A]n ongoing process that creates a chain of feedback loops, with each one influencing the following one… [such that] if one looks at only one of those loops, it is indeed the case that the underlying intuitive process is prior to subjects' conscious reasoning: for each loop at a time, the automatic intuition comes first. But if one steps back and takes a look at the whole chain of feedback loops, what used to look like *idle* confabulation suddenly starts to look like an extremely efficient way of managing one's intuitions.'[151]

On this view of moral intuition and education, intuition itself may be regarded as heuristic, comprised of rules and principles to which structures in the brain are innately attuned, and which continuously updates throughout life through *ex ante* and *ex post* processes of moral education. Notwithstanding the fact that moral judgments are provided automatically and intuitively, therefore, those intuitions nonetheless reflect *rational* moral reasoning, provided that the individual has received appropriate exposure and moral education throughout development and into later life. Equally, those intuitions can become irrational where such moral education is deficient. Two points emerge of critical importance to the present thesis, however: first, the fact that a moral decision-making

[150] Kennet and Fine (2009); citing Fritz Strack and Roland Deutsch, 'Reflective and impulsive determinants of social behavior' (2004) 8(3) *Personality and Social Psychology Review* 220; Russell H. Fazio and Michael A. Olson, 'Implicit measures in social cognition research: Their meaning and use' (2003) 54(1) *Annual Review of Psychology* 297.

[151] Sauer (2012), 271 (emphasis added).

system is automatic and intuitive does not preclude it from being rational and, second, nor does it preclude that system from being updated – *i.e.,* responding to reason.[152]

It is further pertinent to note that this view of intuitive moral decision-making and its interaction with conscious reasoning is entirely commensurate with the role that consciousness is hypothesised to play in decision-making generally throughout this thesis. The evidence considered throughout chapters three to seven of the present thesis strongly suggests that each component of a decision – *what, how, when, whether* and *why* – is first decided unconsciously before reaching the level of conscious awareness. Further, from the general proposition that conscious thought itself does not emerge from a vacuum as a *causa sui* of decisions and action (unless mind / body dualism is accepted and macroscopic physical determinism is denied), it follows that each conscious thought must be preceded by unconscious activity in the brain. When conscious deliberation therefore takes place – such as through a process of *ex post* moral reflection and reasoning – each thought within a chain of reasoning is itself the result of unconscious and automatic processes, and feeds back into those processes in order to produce the next thought.

It is submitted that, what is gained through the process of conscious deliberation is the greater time and mental resources required for any given decision to evolve more fully. For example, whereas an individual might act on their initial, automatic moral intuition which is influenced by racial bias, it requires additional time and mental resources to recognise that bias, to motivate away therefrom, and to engage executive functions in order to overcome the bias. In this regard, it is hypothesised that conscious thought cannot itself directly control or override the underlying automatic processes which give rise to that conscious bias in the first place; but the process of conscious deliberation can provide more time and dedicate greater mental resources towards the requisite underlying unconscious processes which then ultimately do override or decide away from the unwanted bias.

[152] See further Peter Railton, 'The affective dog and its rational tale: Intuition and attunement' (2014) 124(4) *Ethics* 813; Peter Railton, 'Moral learning: Conceptual foundations and normative relevance' (2017) 167 *Cognition* 172.

### 7.2.3. *The Legal Relevance of Intuitionist Models of Moral Decision-Making*

Both the social intuitionist model of moral decision-making and the theory of universal moral grammar carry similar implications vis-à-vis the subjective access available to *genuine* reasons for decisions, and the ability for people to distinguish such genuine reasons from those that are confabulated. The implication from both theories, therefore, is that people's subjective account of their reasons for acting are unreliable at best, notwithstanding the additional difficulty for third parties (such as a jury or judge) to objectively discern whether the explanations offered by a defendant or witness are genuine, confabulated, or outright lies. In this respect, relying upon proof of an individual's subjective mindset – which cannot be accessed objectively and neither entirely trusted when subjectively recounted even by the most honest individual – is an ultimately *un*reliable and *un*safe means of attributing responsibilty for people's decisions and actions.

That notwithstanding, both the social intuitionist model and UMG contribute important offerings to the concept of volition, and the capacity for people to recognise and apply reason to their decision-making. In particular, the previous discussion offers an explanation of how automatic and intuitive decision-making processes can be nonetheless *rational* and, crucially, responsive to reason (this itself arguably being a hallmark of rational thought). In the first instance, it is proposed that any such innate structures that exist in the brain and are amenable to learning moral grammar must nevertheless be exposed to a moral education, just as the linguistic structures in the brain must be exposed to language in the first place in order to learn and later produce language. In this regard, many of the moral lessons that children learn are inherently rational, such as the rules and principles suggested by Mikhail. What is more, it is arguable that the structures in place for learning (rational) moral grammar may arise from biological determinants, such as the principle of self-preservation drawing from the biological flight or fight response, or the prohibition against battery drawing from the biological stress response.

In the second instance, the resulting automatic and intuitive moral decision-making processes in the brain are evidently amendable to a moral education that continues throughout life. Thus, even where the systems and processes involved operate

automatically and beneath the level of consciousness, they still incorporate and respond to reason. Sauer writes:

> '[A]lthough we only rarely have time to reflect about what to do on a given occasion, we did have time to acquire a repertoire of intuitions about what is morally acceptable, over the course of our moral education, which we can produce automatically without any thought. And we do have time to reflect and reason about those intuitions when we are confronted with a special reason to do so – a conversation we had, a new piece of information we gathered, an argument we embarked upon with a friend, or a moral conflict we encountered. This… is where reason comes into play in the production of moral judgment. From the back: because reasoning figures in the acquisition, formation, and maintenance of our moral intuition. From the front: because these moral intuitions are amendable to reflection, once the need for an intermittent episode of moral reasoning has arisen.'[153]

This conclusion is vital to the legal concept of volition, explored more fully in sections 8.1 and 9.3 of the present thesis. In brief, volition consists of two concepts – that people have a capacity to exercise conscious control over their actions, and that their decision-making is *responsive to reason*. This latter capacity appears *prima facie* to be called into question by the body of evidence suggesting that decision-making occurs through automatic, intuitive and unconscious process; if this is the case, how can people make rational decisions and, critically, recognise and respond to good and bad reasons for different decisions and actions?

The answer lies in the preceding conclusions, above. The moral decision-making structures of the brain are capable of recognising reasons because they are formed upon a set of rules and principles that are learned throughout moral education, and in particular during early development. Furthermore, those structures are capable of responding to reason because that process of learning does not terminate upon maturity. Despite being

[153] Hanno Sauer, *Moral Judgments as Educated Intuitions* (Massachusetts Institute of Technology Press 2017), 11.

automatic, intuitive and unconscious, moral intuitions are continually updated by experiences and moral education throughout life such that, when recognisably good (or bad) reasons for action are presented, those reasons impact upon and are incorporated into the underlying decision-making processes. The fact that those processes are automatic, intuitive and unconscious precludes neither their ability to produce rational decisions and responses to different situations, nor their ability to be updated or, in more legal parlance, to respond to reason.

## *7.3. Confabulation and Post-hoc Rationalisation*

Whereas the early split-brain experiments demonstrated the ready ability for non-neurotypical people to confabulate reasons for their decisions, more modern research has continued to reveal how neurotypical people can also (and often frequently) confabulate. Confabulation was itself originally understood as a clinical condition relating to a disorder of memory and delusions;[154] however, clearly confabulation is no longer limited to solely clinical cases and, indeed, 'there may be very little *observable* difference between confabulation and explanation.'[155] This, alongside other evidence explored below, has led to theories suggesting that the verbal reasons people are able to give for their decisions and actions are in fact constructed *post hoc* by the brain as a means of explaining behaviour within a social context, rather than explicit reasons first being generated which then lead to a particular judgment or decision.

### *7.3.1. Confabulation in Non-Clinical Cases*

A key feature of confabulation which makes it difficult to distinguish from explanation is that confabulations are given as genuinely believed reasons with no intention to deceive. As Dennett writes, 'it is not that [people] lie in the experimental situation, but that they

[154] See Hirstein (2005); Martha Turner and Max Coltheart, 'Confabulation and delusion: a common monitoring framework' (2010) 15(1) *Cognitive Neuropsychology* 346.

[155] Ana P. Gantman, Marieke A. Adriaanse, Peter M. Gollwitzer and Gabriele Oettingen, 'Why did I do that? Explaining actions activated outside of awareness' (2017) 24(5) *Psychonomic Bulletin & Review* 1563, 1563 (emphasis added); citing Nisbett and Wilson 1977; Petter Johansson, Lars Hall, Sverker Sikström, Betty Tärning and Andreas Lind, 'How something can be said about telling more than we can know: on choice blindness and introspection' (2006) 15(4) *Consciousness and Cognition* 673.

confabulate; they make up likely sounding tales without realizing they are doing it; they fill in the gaps, guess, speculate, mistake theorizing for observing.'[156] It has been proposed that filling explanatory gaps in the reasons behind our decisions restores a sense of 'agentic coherence and consistency.[157] One effect of this, however, is that individuals themselves are virtually unable to distinguish between when they are confabulating and when they are offering a genuine explanation for their decisions and actions.

In some of the earliest non-clinical experiments by Nisbett and Wilson,[158] subjects were presented with a rack of stockings from which to select their preference as if in a consumer study. Despite all of the stockings being identical, participants overwhelmingly selected from the rightmost pair. However, when asked to explain their choice, subjects never gave the position of the stockings as a reason, some even refuting this outright as a possible explanation. When asked directly about the positioning of the stockings, subjects denied that it had any effect and stated 'either that they had misunderstood the question or were dealing with a madman.'[159] In a similarly simple experiment,[160] subjects waited in a room where they could eat from bowls of "goldfish crackers" and "animal crackers", whilst a confederate also waited in the room and ate from only one of the two available snacks. Subjects readily mimicked the confederate and ate more of the same snack; however, they remained unaware that they were mimicking the confederate and instead reported a subjective preference for the particular snack that they had eaten.

Bar-Anan, Wilson and Hassin conducted four studies to explore confabulated self-knowledge in response to automatic behaviour,[161] which follows in many ways from the studies regarding priming of automatic behaviour considered in sections 3.1 of this thesis,

---

[156] Daniel C. Dennett, 'How to study human consciousness empirically or nothing comes to mind' (1982) 53(2) *Matters of the Mind* 159, 173.
[157] Gantman, Adriaanse, Gollwitzer and Oettingen (2017), 1564; citing Jeffrey W. Cooney and Michael S. Gazzaniga, 'Neurological disorders and the structure of human consciousness' (2003) 7(4) *Trends in Cognitive Sciences* 161; Daniel C. Dennett, 'The self as a responding – and responsible – artifact' (2003) 1001(1) *Annals of the New York Academy of Sciences* 39.
[158] Nisbett and Wilson (1977).
[159] *Ibid*., 244.
[160] Robin J. Tanner, Rosellina Ferraro, Tanya L. Chartrand, James F. Bettman and Rick van Baaren, 'Of chameleons and consumption: The impact of mimicry on choice and preferences' (2008) 34(6) *Journal of Consumer Research* 754.
[161] Yoav Bar-Anan, Timothy D. Wilson and Ran R. Hassin, 'Inaccurate self-knowledge formation as a result of automatic behavior' (2010) 46(6) *Journal of Experimental Social Psychology* 884.

above. Previous research has shown in particular that goals can both be activated and induce related behaviours outside of people's conscious awareness.[162] Furthermore, evidence suggests that primed participants 'generally do not attribute their behavior to the priming manipulation',[163] mirroring the comments by Smeesters, Wheeler and Kay[164] discussed in section 7.1.3, above. Bar-Anan, Wilson and Hassin primed subjects with different goals, namely opposite-sex affiliation, helping others, or earning money, and then asked subjects to choose between two alternatives, one of which could further the attainment of the primed goal. For example, male subjects might be primed to affiliate with a member of the opposite sex, before being offered the choice between two courses, one delivered by a man and the other by a woman.

Across the four studies, subjects remained unaware of the primed goal and its impact on their choices, which significantly followed the prime. Thus, subjects primed to affiliate with the opposite sex generally chose a course delivered by the opposite sex; subjects primed to be helpful to others preferred to play a cooperative rather than competitive game; and subjects primed to earn money preferred to play a trivia game with images of US presidents from currency.[165] Crucially, however, subjects 'failed to identify the extent to which a primed goal influenced a choice and attributed that choice to preferences and dispositions unrelated to the goal.'[166] For example, men who selected a course delivered by a female tutor often cited the course content as the reason for their choice; similarly, subjects who selected a particular game later cited such reasons as their preference for playing that game. Recalling that social behaviours may be automatically activated outside of conscious awareness, Bar-Anan, Wilson and Hassin conclude from these experiments that self-knowledge may be readily prone to error. Moreover, the research

[162] Bargh, Lee-Chai, Barndollar, Gollwitzer and Trötschel (2001); Henk Aarts, Peter M. Gollwitzer and Ran R. Hassin, 'Goal contagion: Perceiving is for pursuing' (2004) 87(1) *Journal of Personality and Social Psychology* 23; Ran R. Hassin, John A. Bargh and Shira Zimerman, 'Automatic and flexible: The case of non-conscious goal pursuit' (2009) 27(1) *Social Cognition* 20.

[163] Bar-Anan, Wilson and Hassin (2010), 885; citing Ayelet Fishbach and Aparna A. Labroo, 'Be better or be merry: How mood affects self-control' (2007) 93(2) *Journal of Personality and Social Psychology* 158; Paschal Sheeran, Thomas L. Webb and Peter M. Gollwitzer, 'The interplay between goal intentions and implementation intentions' (2005) 31(1) *Personality and Social Psychology Bulletin* 87; Azim F. Shariff and Ara Norenzayan, 'God is watching you: Priming god concepts increases prosocial behavior in an anonymous economic game' (2007) 18(9) *Psychological Science* 803.

[164] Smeesters, Wheeler and Kay (2010), 307.

[165] Bar-Anan, Wilson and Hassin (2010), 886 – 892.

[166] *Ibid*., 892.

suggests that 'even when people's behavior is the result of a high-level mental process, such as the goal to help someone, people are often in "the same position as an outside observer" in understanding why they did what they did.'[167]

Other studies exploring confabulation in the context of decisions and behaviour activated outside of conscious awareness have focused on the 'psychological consequences of acting without having an accessible explanation for one's own behavior, or, in other words, of "acting in an explanatory vacuum".'[168] The study by Bar-Anan, Wilson and Hassin, discussed above, provides evidence for provoked confabulation arising in response to being probed about one's behaviour, whilst a number of further studies demonstrate spontaneous confabulation arising, for example, as a result of experiencing a negative affect as a result of the explanatory vacuum.[169] Gantman, Adriaanse, Gollwitzer and Oettingen thus describe how non-clinical confabulation may be 'likened to the way the brain fills in blind spots to create a unified visual field. Specifically, confabulation aims to create a unified image of conscious life without gaps in memory or agentic coherence.'[170]

*

A significant body of the most current research led by Petter Johansson and Lars Hall explores the phenomenon of choice blindness, which refers to the failure by people to notice 'conspicuous mismatches between their intended choice and the outcome they [are] presented with.'[171] In an original paradigm, Johansson, Hall, Sikström and Olsson

[167] *Ibid*; citing Daryl J. Bem, 'Self-perception theory' in Berkowtiz L. (ed.), *Advances in Experimental Social Psychology* (Academic Press 1972), 2.
[168] Gantman, Adriaanse, Gollwitzer and Oettingen (2017), 1565; citing Gabriele Oettingen, Heidi Grant, Pamela K. Smith, Mary Skinner and Peter M. Gollwitzer, 'Nonconscious goal pursuit: Acting in an explanatory vacuum' (2006) 42(5) *Journal of Experimental Social Psychology* 668.
[169] Oettingen, Grant, Smith, Skinner and Gollwitzer (2006); Elizabeth J. Parks-Stamm, Gabriele Oettingen and Peter M. Gollwitzer, 'Making sense of one's actions in an explanatory vacuum: The interpretation of nonconscious goal striving' (2010) 46(3) *Journal of Experimental Social Psychology* 531; Marieke A. Adriaanse, Jonas Weijers, Denise T. D. de Ridder, Jessie de Witt Huberts and Catherine Evers, 'Confabulating reasons for behaving bad: The psychological consequences of unconsciously activated behaviour that violates one's standards' (2014) 44(3) *Journal of Social Psychology* 255.
[170] Gantman, Adriaanse, Gollwitzer and Oettingen (2017), 1570.
[171] Petter Johansson, Lars Hall, Sverker Sikström and Andreas Olsson, 'Failure to detect mismatches between intention and outcome in a simple decision task' (2005) 310(5745) *Science* 116, 116.

presented subjects with pairs of faces between which they were instructed to choose their preference on the basis of attractiveness. On some trials, the subjects were asked to provide reasons for their particular choices. On other trials, the presentation of the subject's "choice" was covertly manipulated such that, in fact, they were being asked to provide reasons for a choice that they had not actually made. Moreover, different time conditions were deployed between the trials such that some face pairs were presented for 2 seconds, some for 5 seconds, and some for as long as the subject required. Whilst most decisions were reached within 2 seconds across all of the time conditions, this design enabled the experiment to compare between those choices that were forced rapidly and those for which the subject had copious time for deliberation.[172]

A number of findings are reasonably surprising; first, subjects detected that their chosen preference had been switched for a different option in only 13% of trials, which increased to 27% on trials where subjects had no restriction on their time to deliberate. These figures held even where the pairs of faces presented bore little resemblance to one another, such that it was 'hard to imagine how a choice between them could be confused.'[173] Furthermore, it would be expected that introspective reports for manipulated and non-manipulated trials would differ, with the former revealing reasons behind a choice whilst the latter being more anomalous in reporting reasons for a choice which had not in fact been made. However, verbal reports were analysed along a number of categories, including length of statements, verb tense used, emotionality, specificity, certainty, and concurrent laughter, with no significant differences found between the reasons provided on manipulated and non-manipulated trials. As Johansson, Hall, Sikström and Olsson write, 'the [manipulated] reports were delivered with the same confidence as the [non-manipulated] ones, and with the same level of detail and emotionality.'[174]

Similar findings have been replicated across a number of scenarios by Johansson and various colleagues, including choice blindness regarding the attractiveness of faces and

[172] *Ibid.,* 117.
[173] *Ibid*.
[174] *Ibid.,* 118.

abstract patterns,[175] consumer choices between different foods, flavours and food ingredients,[176] political preferences between liberals and conservatives,[177] analysing arguments in moral dilemmas and other reasoning problems,[178] and in relation to choices taken online within a virtual world.[179] What is more, there is evidence that the phenomenon of choice blindness can continue to influence and even change people's attitudes over the longer term. In the studies concerning political attitudes, for example, subjects first completed a questionnaire or otherwise gave their political opinions on a number of issues. Some of these responses were then manipulated to present the contrary view and subjects were asked to verify the manipulated responses and sometimes provide the underlying arguments supporting that manipulated response. One-third to one-half of manipulated responses were corrected by subjects across three different studies exploring political choice blindness;[180] however, the third experiment goes further by re-testing subjects' attitudes later during the experiment and again after one week.

Strandberg, Sivén, Hall, Johansson and Pärnamets found that subjects' responses were 'strongly affected by the false feedback' both directly after the experiment and one week later, whilst 'attitude change was much larger if participants were asked to reason about why they had stated the attitude falsely presented as their own compared with when only

[175] Petter Johansson, Lars Hall and Sverker Sikström, 'From change blindness to choice blindness' (2008) 51(2) *Psychologia* 142.
[176] Tracey T. L. Cheung, Astrid F. Junghans, Garmt B. Dijskterhuis, Floor M. Kroese, Petter Johansson, Lars Hall and Denise T. D. de Ridder, 'Consumers' choice-blindness to ingredient information' (2016) 106 *Appetite* 2; Lars Hall, Petter Johansson, Betty Tärning, Sverker Sikström and Thérèse Deutgen, 'Magic at the marketplace: Choice blindness for the taste of jam and the smell of tea' (2010) 117(1) *Cognition* 54.
[177] Lars Hall, Thomas Strandberg, Philip Pärnamets, Andreas Lind, Better Tärning and Petter Johansson, 'How the polls can be both spot on and dead wrong: Using choice blindness to shift political attitudes and voter intentions' (2013) 8(4) *PLoS ONE* e60554; Thomas Strandberg, Jay A. Olson, Lars Hall, Andy Woods and Petter Johansson, 'Depolarizing American voters: Democrats and Republicans are equally susceptible to false attitude feedback' (2020) 15(2) *PLoS ONE* e0226799.
[178] Lars Hall, Petter Johansson and Thomas Strandberg, 'Lifting the veil of morality: Choice blindness and attitude reversals on a self-transforming survey' (2012) 7(9) *PLoS ONE* e45457; Emmanuel Trouche, Petter Johansson, Lars Hall and Hugo Mercier, 'The selective laziness of reasoning' (2015) 40(8) *Cognitive Science* 2122.
[179] Petter Johansson, Lars Hall, Agenta Gulz, Magnus Haake and Katsumi Watanabe, 'Choice blindness and trust in the virtual world' (2007) 107(60) *Technical Report of IEICE: HIP* 83.
[180] Hall, Strandberg, Pärnamets, Lind, Tärning and Johansson (2013); Strandberg, Olson, Hall, Woods and Johansson (2020); Thomas Strandberg, David Sivén, Lars Hall, Petter Johansson and Philip Pärnamets, 'False beliefs and confabulation can lead to lasting changes in political attitudes' (2018) 147(9) *Journal of Experimental Psychology* 1382.

acknowledging its position.'[181] This concurs with further studies demonstrating how preferences can be changed following false feedback procedures.[182] The authors hypothesise that the process of confabulating reasons for a manipulated choice may be what influences people's future attitudes and responses. As they explain, the increase in the average ratings provided by subjects was around 50% when subjects confabulated a reason for their manipulated choice as compared with when they merely acknowledged that the choice was their own, whilst this increase in ratings became twice as large one week later, representing a considerable effect. The authors suggest that this 'shows how the perception and verbalization of one's own reasoning can influence one's attitudes.'[183]

Discussing the research generally, Hall and Petersson write that choice blindness 'drive[s] a large wedge between intentions and actions in the mind' as subjects significantly give 'verbal explanations about choices they never made.'[184] They highlight how the effects of choice blindness have been demonstrated not only in rapid snap decisions but also in decision taken over extended periods of time, meanwhile between 80% and 90% of subjects across studies consistently believed that they 'would have noticed that something was wrong.'[185] Although both significant and surprising, the fact that a majority of subjects generally failed to spot their own manipulated choices is not key to the theme of the present chapter of the thesis. Rather, it is the 'robust, replicable, and often dramatic effect' of subjects unwaveringly providing 'introspectively derived' confabulated reasons for choices that they never made that is the crucial point.[186] Choice blindness, once again,

---

[181] Strandberg, Sivén, Hall, Johansson and Pärnamets (2018), 1393.

[182] Tali Sharot, Stephen M. Felming, Xiaoyu Yu, Raphael Koster and Raymond J. Dolan, 'Is choice-induced preference change long lasting?' (2012) 23(10) *Psychological Science* 1123; Petter Johansson, Lars Hall, Betty Tärning, Sverker Sikström and Nick Chater, 'Choice blindness and preference change: You will like this paper better if you (believe you) chose to read it!' (2014) 27(3) *Journal of Behavioral Decision Making* 281; Keise Izuma, Shyam Akula, Kou Murayama, Daw-An Wu, Marco Iacoboni and Ralph Adolphs, 'A causal role for posterior medial frontal cortex in choice-induced preference change' (2015) 35(8) *Journal of Neuroscience* 3598.

[183] Strandberg, Sivén, Hall, Johansson and Pärnamets (2018), 1394; citing Jamie Barden and Zakary L. Tormala, 'Elaboration and attitude strength: The new meta-cognitive perspective' (2014) 8(1) *Social and Personality Psychology Compass* 17; Zakary L. Tormala and Richard E. Petty, 'What doesn't kill me makes me stronger: The effects of resisting persuasion on attitude certainty' (2002) 83(6) *Journal of Personality and Social Psychology* 1298.

[184] Lars Hall and Petter Johansson, 'Choice blindness: You don't know what you want' (2009) 2704 *New Scientist* 26, 26.

[185] *Ibid*., 27.

[186] Petter Johansson, Lars Hall and Nick Chater, 'Preference change through choice' in Dolan R. and Sharot T. (eds.), *Neuroscience of Preference and Choice* (Elsevier Academic Press 2011), 126.

reveals a disconnect between our decisions and actions on the one hand and the subjectively accessible reasons for those actions on the other. If reasons motivate decision outcomes by building upon first principles, it is inexplicable that people will both readily adopt a choice that they haven't made and then proceed to provide reasons for that choice as if they were genuine and certain. The genuine choice that people made should be underpinned by prior existing reasons which would conflict with a manipulated choice that is presented back and preclude the adoption of new reasons to support that manipulated choice. Rather, choice blindness is more readily explicable if decisions are reached first, following which the brain produces explanations (or confabulations) for those decisions.

### *7.3.2. Post-hoc Rationalisation*

The combination of an apparently poor subjective access to the higher mental processes that give rise to preferences, decisions and behaviour, in conjunction with a ready ability to confabulate reasons for our actions which are seemingly indistinguishable from genuine explanation, leads naturally to the question of why this state of affairs exists? What purpose or function does it serve? Logic dictates that our reasons underpin our decisions whilst evidence suggests instead that decisions precede reasoning; what purpose, then, is served by having access to reasons that may not genuinely correspond to our decisions? A widely held conclusion within academia is that:

> '[M]ost explicit practical reasoning and justifications we offer to others or ourselves are rationalizations, and we instead act on instincts, inclinations, stereotypes, emotions, neurobiology, habits, reactions, evolutionary pressures, unexamined principles, or justifications other than the ones we think we're acting on. Then – and this is the crucial part of the claim – we tell a *post hoc* story to justify the actions that are better explained in these alternative ways.'[187]

[187] Jesse S. Summers, '*Post hoc ergo propter hoc*: some benefits of rationalization' (2017) 20(1) *Philosophical Explorations* 21, 22; citing Fiery Cushman and Joshua D. Greene, 'The philosopher in the theater' in Mikulincer M. and Shaver P. R. (eds.), *The Social Psychology of Morality: Exploring the Causes*

A commonly used analogy is that the process of reasoning acts like a lawyer, providing the best available arguments to explain or justify a particular decision or action. As such, it does not *necessarily* require that those reasons correspond to the genuine reasons behind an action (although the possibility remains open that they may so correspond). Crucially, however, the process of reasoning is not acting as a navigator, first providing reasons upon which to then base the decision of what route to take. This explanation accounts for many of the phenomena considered in this present chapter of the thesis; the brain's ready ability to confabulate exists in order to provide adequate or even persuasive reasons for decisions and actions, not necessarily genuine ones. The experience of being morally dumbfounded suggests that moral judgments are made intuitively and justifications provided second, with moral judgments remaining steadfast even in the absence of suitable justification. And, again, choice blindness demonstrates the disconnection between reasons and decisions, whereby people whose choices have been manipulated nonetheless readily provide arguments to support a choice that they never made, just like the lawyer advocating *post hoc* for the actions of their client.

It is crucial to note that the claim is not made that explicit reasoning *cannot* therefore provide the best explanation for actions. For example, one may reason about how much money they need to withdraw from an ATM, and this reason would likely provide the best explanation for the amount of money that they then proceeded to withdraw. However, 'even if not all reasoning is rationalization, the research shows that we rationalize far more than sincere introspection reveals,'[188] whilst we remain virtually incapable of subjectively discerning the difference between those subjectively accessible reasons which are the genuine causes behind our decisions and those which are confabulated or rationalised *post hoc*. This accounts for a number of further phenomena in cognition and

*of Good and Evil* (American Psychological Association Press 2011); Joshua D. Greene, 'The secret joke of Kant's soul' in Sinnott-Armstrong W. (ed.), *Moral Psychology Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development* (Massachusetts Institute of Technology Pres 2008); Michael S. Gazzaniga, *Who's in Charge? Free Will and the Science of the Brain* (Robinson 2012); Haidt (2001); Benjamin Libet, 'Do we have free will?' (1999) 6(8-9) *Journal of Consciousness Studies* 47; Daniel M. Wegner and Thalia Wheatley, 'Apparent mental causation: Sources of the experience of will' (1999) 54(7) *American Psychologist* 480; Nisbett and Wilson (1977).

[188] Summers (2017), 26; citing Alfred Mele, 'Unconscious decisions and free will' (2013) 26(6) *Philosophical Psychology* 777; Darcia Narvaez, 'The social intuitionist model: Some counter-intuitions' in Sinnott-Armstrong W. (ed.), *Moral Psychology Volume 2: The Cognitive Science of Morality: Intuition and Diversity* (Massachusetts Institute of Technology Press 2008).

decision-making generally, including a range of cognitive biases such as anchoring, belief, and confirmation biases which concern the way people prefer (or are biased towards) certain types of information,[189] mistaken and flawed uses of probabilistic reasoning,[190] and generally rather poor reasoning abilities on simple logical tasks;[191] because the reasoning process is seeking the most appropriate or persuasive reason, and not necessarily the correct or genuine one.

Mercier and Sperber offer an "Argumentative Theory" to account for why a process of *post hoc* rationalisation may have emerged as the dominant means of subjectively explaining our behaviour.[192] To begin on the one hand, the authors distinguish between processes of inference that are unconscious and intuitive, and the 'representational output which necessarily or probabilistically follows from its representational input.'[193] Thus, people may be 'aware of having reached a certain conclusion – be aware, that is, of the output of an inferential process – but… they are never aware of the process itself,' such that inference processes produce *intuitive beliefs* that are 'held without awareness of reasons to hold them.'[194] On the other hand are *reflective beliefs* that are 'held with awareness of one's reasons to hold them;' for example, a reflective belief may be based on trust in its source (*e.g.* a professor, doctor or lawyer), or based on the content of the belief itself, such as its consistency with previously held beliefs. What characterises reasoning, therefore, is an awareness 'not just of a conclusion but of an argument that justifies accepting that conclusion.'[195] However, Mercier and Sperber suggest that 'arguments exploited in reasoning are the output of an intuitive inferential mechanism [and,] like all other inferential mechanisms, its processes are unconscious… and its

[189] Daniel Kahneman, Paul Slovic and Amos Tversky (eds.), *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press 1982).
[190] Daniel Kahneman and Amos Tversky, 'Subjective probability: A judgment of representativeness' (1972) 3(3) *Cognitive Psychology* 430; Amos Tversky and Daniel Kahneman, 'Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment' (1983) 90(4) *Psychological Review* 293.
[191] Jonathan St. B. T. Evans, 'Logic and human reasoning: An assessment of the deduction paradigm' (2002) 128(6) *Psychological Bulletin* 978.
[192] Hugo Mercier and Dan Sperber, 'Why do humans reason? Arguments for an argumentative theory' (2011) 34(2) *Behavioral and Brain Sciences* 57.
[193] *Ibid*., 58.
[194] *Ibid*.
[195] *Ibid*.

conclusions are intuitive.'[196] These intuitive conclusions are about 'representations of relationships between premises and conclusions,' *i.e.* arguments.[197]

Mercier and Sperber continue to explain that the possession of intuitions about arguments exerts an evaluative effect; some arguments are regarded as being stronger and others weaker, and we may have an intuitive preference where arguments compete for opposite conclusions, but these evaluations and preferences are ultimately formed on the basis of unconsciously generated intuitions. This may be understood in relation to the competing neural networks theories presented in section 2.3 of this thesis, whereby different options (such as two or more arguments) are represented by neural networks which compete to reach a threshold at which point a decision (or preference) is reached. Thus, in relation to argumentation, competing arguments are represented in the brain by competing neural networks, and the network that recruits the most evidence or valence represents the argument that is preferred by the individual. Crucially, however, this process of competition occurs outside of conscious awareness with only its conclusion (*i.e.,* the winning argument) reaching consciousness; as such, the conclusions that it produces are rightly categorised as being intuitive. The process of evaluating, accepting and applying the conclusions of arguments is what is commonly referred to as reasoning.

Mercier and Sperber propose that the very function of reasoning lies in relation to human communication and not necessarily, therefore, in order to provide accurate subjective access to the genuine reasons underlying our decisions.[198] "Function" is here understood in the biological sense of the effect of a trait (*i.e.,* reasoning) that 'causally explains its having evolved and persisted in a population.'[199] Although the ability to reason may confer further advantages, it is proposed that it is best adapted for use in argumentation, which would thus be regarded as its main function. Mercier and Sperber rebut potentially competing main functions of reasoning. For example, it has been proposed that reasoning

[196] *Ibid*; citing Philip Johnson-Laird, *How We Reason* (Oxford University Press 2006), 53; Ray Jackendoff, 'How language helps us think' (1996) 4(1) *Pragmatics & Cognition* 1.
[197] Mercier and Sperber (2011), 58.
[198] *Ibid*., 59.
[199] *Ibid*; citing Colin Allen, Marc Bekoff and George V. Lauder, *Nature's Purposes: Analyses of Function and Design in Biology* (Massachusetts Institute of Technology Press 1998).

functions to correct intuition;[200] however, reasoning itself may potentially be a source of new mistakes and, furthermore, tends to rationalise rather than correct intuitive inferences.[201] Conscious reasoning has also been hypothesised as facilitating the ability to deal with novelty and anticipate future events;[202] however, this may be argued to be less of a characterisation of reasoning *per se* and more a quality of cognition and learning in general, the latter of which may be defined as the 'process by which we become able to use past and current events to predict what the future holds.'[203]

Supporting the main function of reasoning as being for communication and, more specifically, argumentation, Mercier and Sperber propose that reasoning 'enables people to exchange arguments that, on the whole, makes communication more reliable and hence more advantageous.'[204] In order to be stable, communication must benefit both senders and receivers, otherwise one or both would cease the exchange. However, such stability is threatened by deceit and dishonesty, and senders may communicate misinformation in order to take advantage of their listener. In order to tackle such misinformation, therefore, people must exercise a degree of 'epistemic vigilance' in order to 'evaluate the communicator and the content of their messages.'[205] Whilst any number of psychological mechanisms might be involved in epistemic vigilance, Mercier and Sperber identify two as being the most important; namely trust calibration and coherence checking.[206] First, trust calibration consists of evaluating different speakers according to their perceived

[200] Daniel Kahneman, 'A perspective on judgment and choice: Mapping bounded rationality' (2003) 58(9) *American Psychologist* 697.
[201] Mercier and Sperber (2011), 59; citing Jonathan St. B. T. Evans and Peter C. Wason, 'Rationalization in a reasoning task' (1976) 67(4) *British Journal of Psychology* 479.
[202] Jonathan St. B. T. Evans and David E. Over, *Rationality and Reasoning* (Psychology Press 1996), 154.
[203] Yael Niv and Geoffrey Shoenbaum, 'Dialogues on prediction errors' (2008) 12(7) *Trends in Cognitive Sciences* 265, 265.
[204] Mercier and Sperber (2011), 60; citing Jean-Louis Dessalles, *Why We Talk: The Evolutionary Origins of Language* (Oxford University Press 2007); Deanna Kuhn, 'Thinking as argument' (1992) 62(2) *Harvard Educational Review* 155; Michael Billig, *Arguing and Thinking: A Rhetorical Approach to Social Psychology* (Cambridge University Press 1996); Dan Sperber, 'Metarepresentations in an evolutionary perspective' in Sperber D. (ed.), *Metarepresentations: A multidisciplinary perspective* (Oxford University Press 2000); Dan Sperber, 'An evolutionary perspective on testimony and argumentation' (2001) 29(1/2) *Philosophical Topics* 401.
[205] Mercier and Sperber (2011), 60; citing Dan Sperber, Fabrice Clément, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi and Deirdre Wilson, 'Epistemic vigilance' (2010) 25(4) *Mind & Language* 359.
[206] Mercier and Sperber (2011), 60.

competence and benevolence,[207] the rudiments of which have been shown to develop in children between three and six years when they learn to distrust malevolent informants.[208] Second, new information must be incorporated with old and, crucially, inconsistencies must be addressed; one strategy may be to dismiss the new information in order to avoid being misled, but if the speaker is assessed as having high competence then it may be more advantageous to accept the new information and revise previous beliefs.[209]

The speaker who wishes to be trusted by their listener(s) may attempt to boost their credibility, but this will not always be possible. Instead, the speaker can offer premises which lead to their conclusions, *i.e.,* they offer arguments, which is itself a use of reasoning. As Mercier and Sperber write, 'reasoning contributes to the effectiveness and reliability of communication by allowing communicators to argue for their claim and by allowing addressees to assess these arguments. It thus increases both in quantity and in epistemic quality the information humans are able to share.'[210] This account of reasoning developing as argumentation in order to facilitate and improve communication concurs with other examples highlighting the importance of sociality in the emergence of other uniquely human cognitive capacities.[211] For example, emphasis has been placed upon the evolutionary role played by the act of cooperating within small groups,[212] for which effective communication plays an obvious and critical role. The development of reasoning through argumentation provided a means of assessing new information, testing competing ideas, and reaching consensus within a social context. Crucially for the purposes of this thesis, however, is that the reasons people give for their actions are

[207] Richard E. Petty and Duane T. Wegener, 'Attitude change: Multiple roles for persuasion variables' in Gilbert D. T., Fiske S. T. and Lindzey G. (eds.), *The Handbook of Social Psychology Vol. 1* (4th ed. McGraw Hill 1998).
[208] Fabrice Clément, 'To trust or not to trust? Children's social epistemology' (2010) 1(4) *Review of Philosophy and Psychology* 531; Olivier Mascaro and Dan Sperber, 'The moral, epistemic, and mindreading components of children's vigilance towards deception' (2009) 112(3) *Cognition* 367.
[209] Mercier and Sperber (2011), 60.
[210] *Ibid*.
[211] *Ibid*; citing Robin I. M. Dunbar, 'The social brain hypothesis' (1998) 6(5) *Evolutionary Anthropology: Issues, News, and Reviews* 178; Robin I. M. Dunbar and Susanne Shultz, 'Evolution in the social brain' (2007) 317(5843) *Science* 1344; Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne and Henrike Moll, 'Understanding and sharing intentions: The origins of cultural cognition' (2005) 28(5) *Behavioral and Brain Sciences* 675.
[212] Benoît Dubreuil, 'Paleolithic public goods games: why human culture and cooperation did not evolve in one step' (2009) 25(1) *Biology & Philosophy* 53; Kim Sterelny, *The Evolved Apprentice: How Evolution Made Humans Unique* (Massachusetts Institute of Technology Press 2012).

produced by a reasoning system that is seeking the most persuasive and convincing argument, and not one that *necessarily* has access to *genuine* reasons for decisions and actions.

Finally, Mercier and Sperber offer a number of predictions derived from their theory which may be tested against existing evidence, some of which has already been considered in this chapter. Thus, they highlight that whereas people may be relatively poor at reasoning on logic puzzles,[213] their reasoning performance becomes significantly better when assessing different *arguments*.[214] In a similar vein, they cite a number of studies where subjects are tested on various verbal or mathematical logic and reasoning tasks both individually and within a group. Performance is predictably stronger within the group setting,[215] where there are many minds contributing and assimilating ideas for the same problem. In one particular type of task, individual performance averaged at a low 10% success,[216] whilst group performance rose significantly to 80%.[217] Most interestingly, however, the evidence suggests that people are generally only willing to change their mind once they have been convinced and, therefore, debate and *argument* has been shown to be essential to improving group performance.[218]

[213] Mercier and Sperber (2011), 61; citing Evans (2002).
[214] *Ibid*; citing Petty and Wegener (1998); Richard E. Petty and John T. Cacioppo, 'Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses' (1979) 37(10) *Journal of Personality and Social Psychology* 1915; Valerie A. Thompson, Jonathan St. B. T. Evans and Simon J. Handley, 'Persuading and dissuading by conditional argument' (2005) 53(2) *Journal of Memory and Language* 238.
[215] Patrick R. Laughlin and Alan L. Ellis, 'Demonstrability and social combination processes on mathematical intellective tasks' (1986) 22(3) *Journal of Experimental Social Psychology* 177; Mark F. Stasson, Tatsuya Kameda, Craig D. Parks, Suzi K. Zimmerman and James H. Davis, 'Effects of assigned group consensus requirement on group problem solving and group members' learning' (1991) 54(1) *Social Psychology Quarterly* 25.
[216] Jonathan ST. B. T. Evans, Stephen E. Newstead and Ruth M. J. Byrne, *Human Reasoning: The Psychology of Deduction* (Psychology Press 1993).
[217] Maria Augustinova, 'Falsification cueing in collective reasoning: Example of the Wason selection task' (2008) 38(5) *European Journal of Social Psychology* 770; Boris Maciejovsky and David V. Budescu, 'Collective induction without cooperation? Learning and knowledge transfer in cooperative groups and competitive auctions' (2007) 92(5) *Journal of Personality and Social Psychology* 854.
[218] David Moshman and Molly Geil, 'Collaborative reasoning: Evidence for collective rationality' (1998) 4(3) *Thinking & Reasoning* 231; Stefan Schulz-Hardt, Felix C. Brodbeck, Andreas Mojzisch, Rudolf Kerschreiter and Dieter Frey, 'Group decision Making in hidden profile situations: Dissent as a facilitator for decision quality' (2006) 91(6) *Journal of Personality and Social Psychology* 1080.

Equally, in relation to the evidence considered throughout this chapter of the thesis, a theory of *post hoc* rationalisation based on providing suitable argument as opposed to genuine reason provides a compelling explanation for each of the phenomena discussed. It accounts for why the brain is so quick and able to confabulate reasons when no other explanation can be found to account for a given judgment, decision or action. It similarly accounts for evidence of a poor underlying subjective access to genuine reasons, although no definitive comment can be made as to how frequently and reliably such introspective assess might be available. Further still, *post hoc* rationalisation accounts for a number of cognitive biases, especially confirmation bias whereby the process of reasoning, acting as an advocate, and focusing on evidence confirms a previously held view.[219]

## 7.4. From Access to Reason and Post-hoc Rationalisation to Legal Responsibility

There are two questions that are virtually inescapable during every criminal trial: *what did you do?* and *why did you do it?* These will typically be used to elicit the *actus reus* and *mens rea* of the offence charged. The prosecution may ask *why* questions in order to try and establish the defendant's criminal state of mind; and, by the same token, the defence may use this line of enquiry to try and negate *mens rea*. The discussion in this chapter of the thesis focuses on this question of *why* people decide and act the way in which they do and, in particular, the source of the reasons that people are able to give. Whilst subjective states of mind currently form one of the core pillars of legal responsibility, it is assumed by law that we not only possess some manner of conscious control over our state of mind but, just as crucially, some degree of introspective access as to what those states of mind actually are. Where the previous chapters five and six of this thesis have called into question the first assumption of *conscious* control over decisions (and, therefore, subjective mental states), the present chapter raises further challenges against the second assumption that we have sufficiently accurate introspective access into our subjective states of mind.

[219] See Mercier and Sperber (2011), 63 – 66.

The experiments discussed in this chapter provide modern and specific examples supporting the original conclusions of Nisbett and Wilson concerning our ability to access the reasons for our actions; '*such introspective access as may exist is not sufficient to produce accurate reports about the role of critical stimuli*.'[220] Contrary to the typical sensation of being able to accurately introspect and justify our decisions with valid reasons, the proposition is that the human capacity to access our *genuine* reasons for decisions is diminished far below the point that subjective experience would lead us to believe; and, indeed, far below the point that the law might similarly assume to exist. Adding to this unreliability in recalling genuine reasons for decisions, people appear to be almost entirely unable to distinguish between genuine explanation and confabulation. When people's reasons are shown to be incompatible with their strongly held opinions and beliefs, there is a tendency for people to continue to rely on their intuitions even in the absence of supporting reason, as in the case of being morally dumbfounded. This further suggests that it is not people's reported reasons that first underpin and inform a decision but, rather, that reasons are constructed *post hoc* in order to rationalise and persuasively justify that decision.

In practical terms, it is virtually unimaginable that inquiries into *what* somebody did and *why* they did it could ever be eliminated from the courtroom and, indeed, the purpose of this thesis is not to suggest that such inquiries are entirely fruitless. Just as consciousness and deliberation have important roles to play in decision-making even if they do not necessarily confer direct *control* thereover, so the inquiry into people's state of mind must continue to offer evidential value in determining the salient facts of any given case, even if it does not necessarily mean that people have any greater subjective access to their reasons. Put differently, just as the inquiry into the reasons of people who are morally dumbfounded can reveal that their beliefs, judgments and opinions (*i.e.,* subjective states of mind) are not supported by sound reason, so a similar inquiry into a person's reasons and motivations can provide the evidence required by the courtroom to adjudicate how likely, logical and compelling those reasons actually are.

[220] Nisbett and Wilson (1977), 246.

What is argued in this thesis, however, is that subjective states of mind are an unreliable basis upon which to rest one of the core limbs of legal responsibility. Notwithstanding the reasons presented in previous chapters of the thesis, the present chapter demonstrates how people appear to have a generally rather poor introspective access into the genuine reasons underlying their decisions, and a ready ability to confabulate reasons without any intention of dishonesty or deceit. Unlike accounts of what somebody has done, which may be corroborated by multiple witnesses, video, photographs and other physical evidence, accounts of a person's subjective state of mind are entirely unverifiable by comparable methods; it is impossible to prove or disprove another individual's conscious experience. To add to this difficulty, the evidence from the present chapter suggests that people are themselves practically unable to subjectively differentiate between genuine explanations for decisions and reasons that have been confabulated; the brain presents the latter with certainty and without hesitation. Because of this, it is virtually impossible to know how often subjectively accessed and reported reasons are genuine explanation or confabulation although, again, the evidence in this chapter tends to suggest that accurate subjective access to genuine reasons is more often poor than it is reliable, and that confabulation may be the norm.

It seems inherently dangerous, therefore, to partially rest the question of legal responsibility upon a factor (*i.e.,* subjective states of mind) to which subjective access is often inaccurate and unreliable, and for which objective verification can never by achieved with any great certainty, and perhaps rather rarely is achieved with accuracy. The second part of this thesis therefore addresses the issues highlighted within the concept of subjective *mens rea*, providing a reformulated conception that is supported by the scientific evidence and subsequently developed theories that have been explored in this first part of the thesis.

# PART TWO

# *8.* Deconstructing *Mens Rea*

> 'The criminal law generally assumes the existence of free will. The law recognises certain exceptions, in the case of the young, those who for any reason are not fully responsible for their actions, and the vulnerable, and it acknowledges situations of duress and necessity, as also of deception and mistake. But, generally speaking, informed adults of sound mind are treated as autonomous beings able to make their own decisions how they will act…'
>
> - House of Lords, 2007.[1]

Criminal offences are generally understood as comprising of three elements which, when satisfied, establish legal responsibility for a criminal act.[2] First, *actus reus* refers to the prohibited criminal act itself, such as killing another person, stealing or damaging another's property *etc.* However, the maxim *actus non facit reum nisi mens sit rea* provides fundamentally that an act 'does not make a man guilty of a crime, unless his mind be also guilty.'[3] This refers to the second element, *mens rea*, which requires that the guilty party held a particular state of mind such as intention, recklessness or dishonesty *etc.,* concurrently with the commission of the *actus reus*. It is the coincidence of *actus reus* with *mens rea* which establishes *prime facie* wrongdoing in the *actus reus*. It is this, for example, which distinguishes between accidentally knocking into somebody and intentionally shoving them; both actions consist of a potential *actus reus* (knocking, hitting or otherwise applying force to another) but only the latter example includes the requisite *mens rea* of intention necessary to establish criminal liability. Finally, there must be an absence of viable defences which, it shall be demonstrated, each relate to a degree

[1] *R v Kennedy* [2007] UKHL 38, [14].
[2] Nicola Monaghan, *Criminal Law Directions* (6th ed, Oxford University Press 2020), 16.
[3] *Haughton v Smith* [1975] AC 476, 491.

of impairment over the defendant's ordinary faculties, voluntariness, and control over actions.


***Fig. l – Elements of Criminal Responsibility.***[4]

There are two broad theories of culpability in academic literature which seek to account for the connection between *mens rea* and responsibility for actions. The "subjective" account asserts that culpability 'depends upon morally defective choices' and is associated with the traditional subjective mental states such as intention, recklessness and knowledge.[5] Furthermore, as may be gleaned from the succinct definition, the subjective account of culpability carries the assumption that people have online, conscious control over their choices and that, as such, it is the fact that a given defendant makes an immoral choice that establishes legal (and moral) culpability for their actions. Conversely, the "objective" account 'grounds fault in *conduct* rather than choices, arguing that an action attracts blame if in inflicts harm when a reasonable person would not have acted that way.'[6] This is more closely associated with crimes of negligence, which do not refer to a subjective state of mind but rather the unreasonable failure to avoid breaching some legal duty which consequently causes harm.

Drawing from the first part of this thesis, the following chapter first deconstructs subjective *mens rea* and its various assumptions, before chapter nine proceeds to reconstruct *mens rea,* advocating for the more objective account. Specifically, under current conceptions, culpability or responsibility for criminal acts (*actus reus*) is established through *mens rea*. *Mens rea* is principally formulated as different subjective states of mind (*e.g.,* intention, recklessness, knowledge, belief, *etc.*), supported by an underlying presumption that all adults possess online, conscious control over their

[4] Monaghan (2020), 17.
[5] Andrew P. Simester, John R. Spencer, Findlay Stark, G. R. Sullivan and Graham J. Virgo, *Simester and Sullivan's Criminal Law: Theory and Doctrine* (7th ed. Hart Publishing 2019), 9.
[6] *Ibid.*

decisions and actions, and are able to recognise and respond to reason when deciding how to act. Because people are presumed to possess such control, the existence of the requisite *mens rea* (in coincidence with the prohibited *actus reus*) denotes moral blameworthiness, justifying the intervention of the criminal justice system. That is to say, the moral defendant ought to have exercised their conscious control so as not to commit a prohibited criminal act with a morally blameworthy state of mind.

The first part of this thesis has presented evidence from neuroscience and psychology which, it is submitted, undermines the presumption of conscious control over decisions and actions, at least to the extent that such self-control is provided by consciousness *per se*. Furthermore, it is submitted that the evidence presented casts serious doubt over the reliability of subjective states of mind as a basis for establishing legal responsibility. The existence of some particular state of mind (such as an intention to commit a criminal act) does not *alone* prove that a conscious, deliberate choice has been made by an individual acting as an agent, as opposed to a decision resulting from automatic and unconscious processes, potentially instigated (primed) by an entirely exogenous source. What is more, it is contended that people have a generally poor ability to subjectively introspect the *genuine* reasons for their actions, whilst the same cannot reliably be discerned objectively through observation. This renders the courtroom inquiry into a person's subjective state of mind at the time of a given offence as being a potentially arbitrary and considerably unreliable factor on which to focus the attribution of legal responsibility for actions.

It is therefore submitted that *mens rea* should neither focus on subjective states of mind, nor be understood as reflecting moral blame. Rather, it is more coherent to understand *mens rea* as denoting unreasonable conduct, committed by an individual who possess the requisite mental capacities necessary to be held responsible for their actions. This admittedly broad and general notion is given specific iteration through the different formulations of *mens rea*; however, it is further submitted that these formulations should follow a hybrid objective / subjective format. Thus, forms of *mens rea* such as intention and recklessness *etc.,* are given entirely objective definitions, but these are applied taking into consideration the specific subjective circumstances of the defendant in every case.

Finally, underlying this hybrid objective / subjective conception of *mens rea* is the assumption that people possess an "*ordinary*" degree of self-control commensurate with the capacity to be responsive to reason, this assumption itself being supported by the scientific evidence. Thus, the criminally responsible defendant is that individual who commits a prohibited *actus reus* with the requisite objective *mens rea* as applied to their particular subjective circumstances, and absent of any defence negating their capacity to appreciate the nature and consequences of their actions, or the assumptions of reasons responsiveness and ordinary control. The guilty defendant's conduct is regarded as reflecting a criminally unreasonable standard of behaviour (as opposed to reflecting moral blame), because any other person in the same subjective circumstances as the defendant would *reasonably* be expected to exercise the requisite reasons responsiveness, ordinary self-control, and awareness of the nature and consequences of their own actions that are sufficient to prevent them from committing the prohibited criminal act. This reconstruction of *mens rea* is extrapolated more fully in chapter nine of this thesis, below.

## 8.1. ***The Assumptions of Conscious Control and Capacity for Reason***

Allen, Derry and Loveless write, it is said that *mens rea* 'consists of a "guilty state of mind"' such as intention, recklessness or dishonesty, which 'represent states of mind where [the defendant] will have decided or chosen to bring about a result prohibited by the criminal law or will, at least, have realised to a greater or lesser extent that the result would happen.'[7] Central to this focus upon what a defendant has decided – *i.e.,* their subjective state of mind – is the presumption that people's actions are voluntary, which is to say that the defendant 'had control over her conduct (act or omission) at the relevant time.'[8] As this is ordinarily presumed in every case, it is principally the role of legal defences raised by the defendant to call into question the voluntariness of a defendant's actions, whether due to the effects of some medical condition (*e.g.,* automatism and insanity), some undue pressure or force from another (*e.g.,* duress and self-defence), or

[7] Janet Loveless, Mischa Allen and Caroline Derry, *Complete Criminal Law: Text, Cases, and Materials* (7th ed. Oxford University Press 2020), 88.
[8] John Child and David Ormerod, *Smith, Hogan, and Ormerod's Essentials of Criminal Law* (3rd ed. Oxford University Press 2019), 91.

arising from the circumstances surrounding the offence (*e.g.,* necessity and involuntary reflex).

A number of prominent jurists and philosophers have considered the potential connections between consciousness in particular and the type of control over behaviours that is deemed necessary for legal and moral responsibility. Morse writes that 'consciousness and action are central to the law's view of the person.'[9] He continues to explain this legal view of the person (which is encapsulated in the presumption of voluntariness) as a 'conscious… creature capable of practical reason, an agent who forms and acts on intentions that are the product of the person's desires and beliefs… [and] can act for and respond to reasons.'[10] Morse's rationale is that both law and morality are action-guiding, which would not be possible unless people could use rules within their practical reasoning. The link with consciousness is an assumption that it is this which enables the online control of behaviour and actions that is, in turn, necessary in order to be responsive to reasons.

In a similar vein, Raz submits that people are responsible 'if and only if they have the capacity for rational action,' which extends beyond simply the faculties of reasoning and decision-making but includes also 'perception, memory, and control of the body without which one cannot act effectively.'[11] He writes further that 'actions are guided by the agents' powers of rational agency when they are performed for, what the agents believe to be, an adequate reason, and their performance is controlled and guided by the agents' beliefs about what reasons they have and what conditions obtain.'[12] Here, again, Raz is associating legal responsibility with conscious reasoning processes; agents are responsible when they 'consciously act in a particular way' or when they act without such awareness within a 'domain of secure competence' within which such conscious control

[9] Stephen J. Morse, 'Determinism and the death of folk psychology: Two challenges to responsibility from neuroscience' (2008) 9(1) *Minnesota Journal of Law, Science & Technology* 1, 4; see also Stephen J. Morse, 'The non-problem of free will in forensic psychiatry and psychology' (2007) 25(2) *Behavioral Sciences & the Law* 203.
[10] Morse (2008), 4 – 5.
[11] Joseph Raz, *From Normativity to Responsibility* (Oxford University Press 2011), 227.
[12] Joseph Raz, 'Responsibility and the negligence standard' (2010) 30(1) *Oxford Journal of Legal Studies* 1, 5.

would ordinarily be expected.[13] Moore and Hurd take these ideas further, focusing in particular on the notion of *conscious* choice underlying responsibility for actions.[14]

A yet stronger claim is made by Levy who presents the "consciousness thesis", stating that 'consciousness of the facts that give our actions their moral significance is a necessary condition for moral responsibility.'[15] Considering work such as from Libet, discussed in chapter five of this thesis, and the broader claim that all conscious experience must be the product of prior unconscious causes argued throughout Part One of this thesis, Levy concedes that much of the processes behind decision-making – assigning value to different options, weighing up those options and reaching a final assessment – may indeed by 'screened off from consciousness.'[16] This is not to say that consciousness cannot be 'casually efficacious' in our decision-making;[17] and, indeed, chapter five to seven of this thesis have hypothesised how consciousness can indeed influence and change the qualitative outcome of decision-making processes by affording more time and mental resources to those processes. In this sense, Levy's claim that consciousness is causally efficacious is uncontentious. However, Levy continues to further assert that consciousness in turn provides the degree of control over behaviour necessary for responsibility, and that an agent must be 'conscious of the moral significance of their action in order to exercise responsibility-level control over it.'[18]

One of the central submissions in this thesis is that any presumption of direct, online, *conscious* control over decisions is currently unsupportable. In particular, section 3.1.2 of this thesis revealed evidence that goals, intentions and other subjective mental states may be primed and processed entirely outside of conscious awareness and, therefore, conscious control. Furthermore, chapter three presented evidence that the outcome of decisions may be accurately predicted from neural activity prior to individuals becoming

[13] See further Nicola Lacey, 'Responsibility without consciousness' (2015) 36(2) *Oxford Journal of Legal Studies* 219, 226 – 227.
[14] Michael S. Moore and Heidi M. Hurd, 'Punishing the awkward, the stupid, the weak, and the selfish: The culpability of negligence' (2011) 5(2) *Criminal Law and Philosophy* 147; see also Michael S. Moore, 'Choice, character, and excuse' (1990) 7(2) *Social Philosophy and Policy* 29.
[15] Neil Levy, *Consciousness and Moral Responsibility* (Oxford University Press 2014), 14.
[16] *Ibid*., 23.
[17] *Ibid*., 24.
[18] *Ibid*., 111.

aware of reaching any decision. This evidence was taken further in chapters five and six of the thesis, which revealed respectively that both decisions to act and decisions to veto a particular action are each determined first by unconscious activity in the brain before arising to the level of conscious awareness. Again, this is argued to preclude *conscious* control over our decisions and actions. Indeed, it is proposed generally that a deterministic account of human behaviour and decision-making necessarily precludes consciousness acting as a homunculus or *causa sui* that is able to directly control decision outcomes independently from any prior unconscious neural activity.

As has been posited across chapters five to seven of this thesis, however, the denial of online conscious control does not mean that that consciousness has no efficacious role to play in decision-making; nor, indeed, does it mean that we do not have some form of control over our actions. Regarding the former claim, the hypothesis is briefly reiterated that consciousness (a) provides greater time and mental resources to a process of deliberation, and (b) may operate as a necessary interface to translate multiple parallel decision-making processes into a single, serial experience for interaction with the world. These hypotheses need not necessarily be accepted, however, for the overall purposes of this thesis. Rather, the critical point is that the evidence suggests that consciousness *per se* does not deliver online control over decisions and actions and, therefore, the presumption in law that we possess such direct conscious control requires modification.

The second claim, above, is that people do of course have some form of control over their behaviour and actions. Such control can operate "online", such as when a sportsman carefully controls the movement of their hands and arms in order to shoot a ball into a target. And, indeed, people exhibit a more general form of self-control over which actions they carry out in the first place, such as when refraining from eating junk food whilst on a diet or resisting buying a new pair of shoes in order to save money. However, once again, the evidence suggests that neither operation of control is the product of consciousness *per se*. Regarding more specific self-control exerted over individual decisions, section 6.2 of this thesis again presented evidence suggesting that veto decisions are initiated unconsciously just like action decisions, leading to the conclusion that it is not consciousness itself which produces the decision to veto an action. Equally,

the discussion in that chapter concerning inhibitory control as an executive function further suggests that such control does not originate in consciousness.

Concerning the online control of motor actions, one of the principal means through which control is exerted over movements in real time is through sensory prediction errors, explored in chapter four and section 4.2 of this thesis in particular. During movement, the motor system is 'issuing commands for movement, and it is also generating predictions of the anticipated sensory consequences of those movements.'[19] The brain then compares the predicted outcome of a movement with its actual outcome informed by sensory and somatosensory feedback, and 'sensory prediction errors occur when the actual feedback doesn't match these predictions.'[20] Thus, when a darts player misses their target or a piano player hears that they have hit the wrong key, this information is used to both adjust ongoing movements and learn for future movements. It is the element of prediction that is particularly crucial here, however, because it can take between 50 to 150 milliseconds for motor signals to be issued from the motor cortex and for sensory signals to be received for the consequences of that action.[21] On the one hand, this time would be too long for the brain to be able to rely solely on feedback in order to correct errors in movements; on the other hand, this time is too quick to reach the 500 milliseconds of stimulation necessary to reach the neuronal adequacy for conscious experience, discussed in section 5.2.1. That is to say, consciousness itself is too slow to facilitate the online control of motor actions through prediction error.

Indeed, it is further submitted that far from consciousness *per se* being the source of self-control that is assumed in law, it is a learned degree of self-control that enables people to consciously deliberate more fully. Considering ***figure d*** in section 2.3.1 of this thesis, a person with low self-control may choose option B when the time reaches the vertical line to the left of the graph. In this instance, the individual's lack of self-control has resulted in them deciding quickly and impatiently, choosing the first preference to reach awareness. However, with greater self-control the individual could refrain from simply going with

[19] Michael S. Gazzaniga, Richard B. Ivry and George R. Mangun, *Cognitive Neuroscience: The Biology of the Mind* (5th ed. W. W. Norton & Co. 2019), 369.
[20] *Ibid*., 369 – 370.
[21] *Ibid*., 370.

their first choice and take more time to consciously consider the options. In this instance, the individual chooses option A, whose neural representation in the brain recruited the greatest evidence and support. Crucially, if conscious deliberation offers greater time and mental resources to a particular decision, a degree of self-control is a *prerequisite* so that a decision is not taken pre-emptively before conscious deliberation has had an opportunity to proceed. From this perspective, it is a degree of executive self-control that is first required in order to facilitate conscious deliberation of a decision, and not the fact of consciousness itself which delivers that necessary self-control.

Consequently, it must be concluded that the presumption of *conscious* control over decisions and actions within the *mens rea* concept of volition must be modified, as it does not concur with the empirical evidence. Nevertheless, what does remains is what might be described simply as an "ordinary" capacity for self-control. That is to say, the neurotypical brain contains automatic mechanisms which operate to ensure that the body physically performs the actions that are instructed in order to implement (unconsciously reached) decisions to act. In this regard, most people commonly experience the ability to accurately reach and grasp a glass as intended, rather than always knocking it over. As was discussed in chapter four of this thesis, the brain creates predictions of the actions that the body is instructed to carry out, and matches sensory and somatosensory feedback with those predictions in order to monitor whether or not an action is being performed correctly, and to provide online corrections in the case that prediction errors arise. What is more, the effects of an impairment of the "ordinary" capacity for self-control are readily visible, such as anarchic and alien hand syndromes, tremors, seizures and sleep-walking, and even obsessive-compulsive disorder and addiction.

In practical legal terms, replacing the presumption of conscious control with a presumption of ordinary self-control will still serve its function within the concept of volition. Specifically, conscious control of actions need not be proven as part of establishing legal responsibility, but may be disproven as part of establishing certain defences such as automatism or loss of control. It is submitted that the presumption of ordinary self-control can fulfil precisely the same function; it will remain unnecessary for the prosecution to prove that any defendant was possessed of their ordinary capacities of

self-control as part of establishing their legal responsibilty. However, the impairment or total absence of the ordinary capacity for controlling bodily actions to conform with intentions can continue to provide the basis of defences which ultimately rest upon a diminution of this capacity, such as automatism, diminished responsibility, insanity and loss of control. Indeed, replacing the presumption of conscious control with that of ordinary control has far greater implications for theories of punishment, considered in chapter twelve of this thesis, than it does for the prior concept of criminal liability.

*

Notably concerning the presumption that people can use and apply reason in their decision-making, section 7.2 of this thesis did present clear evidence of people's general ability to reason. Going further, section 7.2.1 presented the theory that reasoning in fact emerged as a means to facilitate reliable communication through argumentation, and it is within such social contexts that the human propensity for effective reasoning may be fully appreciated. Further, section 7.2.2 discussed the various ways in which even automatic and unconscious processes can incorporate reasons into decision-making. Therefore, the second presumption within the concept of volition – that people generally possess the capacity to recognise and respond to reason in their decision-making – remains relatively supported in the neuropsychological research. On the one hand, chapter seven of this thesis proposes that people do not generally form decisions in a bottom-up process, starting with first principles and building thereupon in order to arrive at final decisions. Rather, it is submitted that reasons for decisions are more typically produced through *post hoc* rationalisation for an otherwise largely intuitionist decision. In other words, the brain reaches decisions through significantly more intuitionist approaches, and then retroactively produces the best arguments or justifications to explain that decision.

On the other hand, chapter seven also proposed that this process of *post hoc* rationalisation evolved as a means to explain decisions and persuade others (through argument) to decide similarly within a social environment. Indeed, the social intuitionist model of moral decision-making proposes that the intuitionist and social elements of decision-making are more influential, albeit allowing for the possibility that reflection, reason and deliberation

can also be engaged to arrive at reasons for decisions, when these faculties are effortfully engaged. Thus, even if a large majority of "online" and "in the moment" decision-making emerges from automatic and intuitionist processes, it is correctly presumed that those processes can and do still recognise and apply reason to decision-making. For example, if a short-tempered man is knocked in a bar and spills his drink, he may assume that he is under some form of attack and respond by turning and raising his fist. However, if in the brief moment before any punch is thrown, the would-be assailant raises his arms, apologises, and makes clear that it was an accident, this is information that the short-tempered man's brain nonetheless receives and processes, perhaps resulting in the automatic decision to veto the violent response. Notwithstanding that the man's decision-making processes are operating automatically, they can nevertheless recognise reason (*i.e.,* that an accident has occurred and no attack is forthcoming) and apply that reason to their decision (*i.e.,* by vetoing the decision to respond aggressively).

By way of further example, most people are familiar with the experience of consciously deliberating some past action or hypothetical, and considering how they might act in a similar situation. Here, again, it is fundamentally proposed that each thought within the deliberative processes that arises to conscious awareness does so as a result of prior unconscious cerebral activity. In this regard, the process of conscious deliberation is more a process of becoming consciously aware of the outputs from otherwise unconscious and automatic decision-making networks in the brain. Nevertheless, here again people are clearly able to apply rules of logic and reason in order to deliberate a particular decision rationally. Again, the fact that such "conscious" deliberative processes in fact result from numerous automatic processes running in parallel, does not necessarily detract from the fact that those automatic processes can apply rules and principles in order to arrive at decisions that are logical and rational. In this regard, the social intuitionist model of moral reasoning emphasises that such conscious deliberation and reflection on moral judgments is generally slow and effortful, such that it is engaged far less commonly than automatic and intuitionist judgment, and it is less influential when moral judgments are being made rapidly "in the moment" of acting.

## *8.2. Subjective Mental States*

Part One of this thesis has laid out a wide range of evidence from neuroscience and psychology which calls into question legal assumptions regarding peoples' capacity to exercise online conscious control over their decision outcomes and to subjectively access genuine reasons for those decisions. Certainly, no single experiment or even collection thereof abrogates entirely either the role of consciousness in deliberation, nor the capacity to apply reason to rational decision-making, both of which are generally accepted as having *some* degree of influence over decision-making processes and outcomes. However, that this should offer any greater degree of *control* over the *outcome* of a particular decision is submitted to be a fallacy, which undermines the assumption in law that people possess a capacity for direct and on-line conscious control over the decisions that they make. To put differently, there is little evidence of it being possible to consciously direct or *force* the automatic and unconscious decision-making networks in the brain to produce a particular outcome.

The implications of this fallacy are most egregious for the concept of *mens rea* and, in particular, the resting of legal responsibility upon proof that certain subjective states of mind coincided with the defendant committing a prohibited criminal act.[22] Referring to blameworthy states of mind such as intention, recklessness and dishonesty, it is submitted that the various discussions across chapters three to six of this thesis undermine proof of subjective mental states as a reliable basis upon which to rest the question of legal responsibility. Taking each component of decision-making with its key criticisms in turn and beginning with the *what* component, the introduction to this thesis opened with the description of a clinical case in which a brain tumour was the undoubted cause of various changes to a patient's behaviour, decision-making and self-control. Extrapolating this analogy to all decisions being the result of prior causes, the *what* component of any decision is similarly and inescapably the result of prior unconscious activity in the brain, which may or may not result in a particular (potentially criminal) decision. This was demonstrated in chapters four and six of this thesis through the phenomenon of priming, which has been shown to exert influential (if subtle) effects over decision-making across

[22] Allen, Derry and Loveless (2020), 88; see also Earl Fruchtman, 'Recklessness and the limits of *mens rea*: Beyond orthodox subjectivism: Part I' (1987a) 29(3) *Criminal Law Quarterly* 315.

a disparate range of topics and areas, and through the ability to accurately predict decision outcomes before people become consciously aware of them.

Of particular note is the potential for goals to be primed, processed, and executed into final motor actions entirely outside of conscious awareness. This could equally occur in relation to a criminal intention or state of recklessness, *etc.* (*i.e.,* subjective *mens rea*), whereby some criminal intent is unconsciously primed, processed and executed into action without or, at least, before the individual becoming consciously aware thereof. In such circumstances it would be natural to question whether the individual can fairly be held responsible for some goal or intent over which they had no conscious awareness or control. Indeed, an analogous treatment may be drawn with legal defences of insanity and automatism, which claim that a defendant acted to various degrees without conscious control over behaviour as a result of some medical condition. If a goal or criminal intention can similarly (and perhaps even regularly) be activated and executed without conscious intervention in neurotypical defendants, the question arises why a medical condition should be sufficient to provide a legal defence when the capacity for conscious control is absent in both the clinical and neurotypical examples, this capacity being the principal assumption underlying subjective *mens rea*.

The discussion of the *how* component of decision-making in chapter four of the thesis reveals possible explanations for the human experience of everyday conscious control over actions, notwithstanding the conclusion that such control must actually be unconscious and automatic, below. Specifically, evidence suggests that the brain predicts the future outcome (and, in particular, value or valence) of planned actions and compares these against the actual outcome of a performed action. The sensation of agency or control over that action is therefore determined by the concurrence of the prediction and feedback for any given action, whilst discrepancies between these factors (*i.e.,* a prediction error) results in a reduced sense of agency. It is hypothesised that this feature enables the brain to learn how to evaluate and control bodily motions during development and learning, and to attribute actions and their consequences to the self, whereby the brain learns to improve the accuracy of purposeful and intentional actions by using feedback from every action to improve the prediction of the outcome of that action in the future. Whilst this

does not necessarily relate directly to *mens rea*, it does provide some explanation for the gap between the subjective experience of agency and conscious control over actions and the likely reality that such control is in fact an automatic and unconscious process.

Regarding the *when* component, the evidence discussed in chapter five suggests that decisions are reached in the brain prior to conscious awareness thereof, further undermining the potential for conscious control over such decisions. Thus, again, if the brain unconsciously forms a criminal intention or other *mens rea* outside of conscious awareness, it is not possible for that decision to be controlled consciously. Further, the evidence equally suggests that the decision as to *when* to initiate a particular action also arises as a result of unconscious activity in the brain of which the individual later becomes conscious. The law emphasises the coincidence of *mens rea* and *actus reus* because it is assumed that a person who intends a particular action or acts with foresight of a particular risk had sufficient conscious control to decide to do otherwise. However, the research indicates that the coincidence of *mens rea* and *actus reus* is just that – a coincidence; the happenstance of whether or not the relevant automatic brain networks produced the requisite excitation to initiate a given action, with sufficient valence and duration to also arise to the level of conscious awareness.

Libet posited the possibility for a conscious veto in the final moments before a decision is executed into motor action; however, the discussion of the *whether* component in chapter six of this thesis revealed that even a conscious decision to veto an action is itself the result of prior unconscious cerebral activity. Furthermore, the same rationale is applicable to decisions that are consciously deliberated over a period of time. Each conscious thought, judgement or evaluation – and, indeed, decision-making strategy or criteria to be followed – must be the product of some prior subconscious neural activity which causes that conscious experience to occur. The denial of this proposition would be to introduce consciousness as an uncaused homunculus in the decision-making process, a position that is largely discredited by the cognitive sciences, philosophical monism, and the presumption against free will which underlies this thesis. Whilst the answer remains unknown, a significant question nonetheless rests over the precise role that consciousness plays in the decision-making process.

This thesis has hypothesised that conscious may provide a necessary interface in order to translate multiple competing parallel processes into a singular experience through which agents can interact serially with the world. It was further posited that the process of conscious deliberation of a decision lends both time and mental resources (*i.e.,* energy, attention, concentration, *etc.*) to that decision-making process. This undoubtedly can change and often improve a decision, which is why conscious deliberation and rational thought is valued in decision-making; yet, this does not necessitate nor denote that conscious deliberation lends any degree of *control* over the outcome of any given decision. Indeed, the evidence from the discussion of the *when* and *whether* components of decision-making suggests that decisions – and, therefore, the formation of any subjective *mens rea* – are first reached by the brain's various automatic decision-making networks, before the result of those networks reaches the level of conscious awareness as a single, unified decision.

This applies also to each conscious thought and conclusion determined through a process of conscious deliberation over time, as well as any final conscious decision to veto a particular action. It proceeds as a matter of logic that, with any conscious decision to act or veto a particular action first being determined before conscious awareness, consciousness itself cannot be the uncaused source of any degree of *control* over a particular decision *outcome*, notwithstanding that it may indirectly influence or change a decision in the manner hypothesised. Indeed, it is posited above that self-control is required first in order to facilitate effective conscious deliberation, and not the reverse, otherwise people would always pre-emptively initiate their first instinctive choices before ever having the opportunity to engage in conscious deliberation.

Taken together, the discussion from chapters three to six of the thesis significantly undermines reliance upon subjective mental states as a safe and reliable means of ascribing criminal responsibility. Not only may such mental states arise as a result of entirely exogenous influences as opposed to expressing an agent's authentic and considered deliberation but, upon such a criminal state of mind arising, there is little evidence that people can *consciously* choose to do otherwise than that which their automatic and unconscious decision-making processes decide.

Finally, the discussion of the *why* component revealed three key findings: first, that people have poor subjective access to the genuine reasons or motivating factors behind particular decisions, but a ready ability to confabulate reasons which subjectively appear to be certain whilst being objectively unconnected to the real reasons behind a decision. It is noted that no concrete claim can be made regarding how often subjective introspection is accurate, however; second, it is virtually impossible for people to subjectively distinguish between genuine reasons for decisions and reasons that have been confabulated, both being experienced with apparent certainty. Thus, the rate of disconnection between genuine reasons for actions and the subjective recall thereof does not need to be too frequent – one-quarter, one-third? – before this becomes a major concern, if the result is that a significant proportion of subjective accounts of reasons for actions cannot be relied upon. Third, it is likely that the verbal brain constructs reasons for our decisions and actions through a process of *post hoc* rationalisation, acting as the lawyer arguing in support of their client after the fact, rather than as a navigator directing a particular course of action. Taken together, this suggests that there may be a serious disconnection between the genuine motivations that underlie any given decision and the reasons that are available to subjective recollection although, again, the latter is not abrogated altogether.

It is often stated that the law is not interested in the motives behind people's decisions – why did the defendant decide to kill their victim – but in their intentions, *i.e., mens rea* or subjective state of mind. As much as this may be true, there are two lines of inquiry in any criminal trial that are arguably of utmost prominence: *what did a person do*? and *why did they do it*? The prosecution will use the latter question with the intention of eliciting some criminal *mens rea*; for example, the defendant who offers the (*prime facie* irrelevant) motive of obtaining an inheritance may unwittingly reveal their intention (*mens rea*) to commit murder. Equally, the defence will use the latter question in an attempt to disprove *mens rea*. Thus, as much as motive *per se* is irrelevant to the question of legal responsibility, inquiries as to *why* particular decisions or actions were taken are ubiquitous throughout criminal trials, the responses to which often form the basis for inferring or negating the requisite *mens rea*. The evidence considered in chapter seven, however, casts serious concerns over the reliability of such subjective recounts of the

reasons underlying decisions and, therefore, the inferences drawn therefrom with regards to a defendant's subjective state of mind.

Once again, no absolutely concrete statement can be made with regarding the prevalence of *any* of the effects and phenomena discussed in the first part of this thesis. However, the weight of the body of evidence raises serious doubts over the reliance on subjective *mens rea* as a key determinant of legal responsibility. Even where the phenomena discussed are relatively uncommon, it further appears to be virtually impossible to differentiate between these effects and their genuine equivalents. That is to say, subjects are virtually unable to differentiate between endogenously generated and externally primed goals (the *what* component); or between genuine and confabulated reasons for decisions. If the subjective experience is blind to these distinctions, it is unlikely that the courtroom can obtain any greater objective insight into that subjective experience.

Returning to the clinical analogy at the introduction to this thesis, just as a brain tumour caused the changes in the patient's decisions and behaviour, so it is posited that it is the result of automatic and unconscious activity amongst the various neural networks engaged in the *what, how, when, whether,* and *why* components of a given decision that result in a decisional outcome. Following from the discussion across Part One of this thesis, there appears to be a limited (if any) scope for conscious control over any of these components to compel the overall outcome of a decision itself and, therefore, limited (if any) control over what subjective state of *mens rea* an individual may find themselves possessing. That is to say, it is unlikely that a defendant can have direct, online, conscious control over their subjective state of mind. Consequently, the law incorrectly assumes the typical capacity for conscious control over decisions and actions, and subjective *mens rea* amounts to an unreliable basis upon which to establish legal responsibility.

## *8.3. Moral Blame*

The final element of *mens rea* requiring review in light of the present thesis is the current understanding that *mens rea* denotes moral blameworthiness onto the actions of a given defendant. However, it is important to note that this link between *mens rea* and morality

has not always existed. Indeed, Gardner identifies at least five broad eras during which the concept has come to represent different things.[23]

The earliest origins of *mens rea* within the common law legal system can be traced to Anglo-Saxon law prior to the Norman conquest of England in 1066. Where one person intentionally or accidentally killed another, they were required to pay compensation ("*wergeld*") to the victim's family otherwise the family could establish a feud against the killer. "Homicide" thus referred to killing in which the killer publicly admitted their action and paid compensation or established a feud with the victim's family. In contrast, "murder" referred to a form of 'dishonourable killing' following which the victim's family were unable to collect compensation or establish a feud, for example, because the killer had concealed their actions, hidden the body, or otherwise failed to publicly admit their actions. The act of killing another was therefore not regarded as an offence against the crown or the public in general, but was an offence 'against the victim and his family' in a manner that is broadly analogous to civil offences against private individuals today.[24] More importantly, the element of "*mens rea*" that distinguished homicide from murder concerned secrecy on behalf of the killer and an inability to obtain restitution on behalf of the victim's family, rather than any moral condemnation of the act of intentional killing itself.

The payment of *wergeld* became "murdrum" following the Norman conquest of 1066, the payment of which remained due for killing another until its abolition in 1340.[25] This same period saw the development of the concept of the "King's peace", which extended from protecting the royal household to refer more widely to the 'normal and general safeguard of public order.' [26] Thus, the offence of killing another (or homicide) transformed from a private wrong into one falling within the jurisdiction of the King's peace. During this same period, canon law grew to exert ever greater influence over the

[23] Martin R. Gardner, 'The *mens rea* enigma: Observations on the role of motive in the criminal law past and present' (1993) 3 *Utah Law Review* 635, 641.
[24] Thomas Benedict Lambert, 'Theft, homicide and crime in late Anglo-Saxon law' (2012) 214(1) *Past & Present* 3, 9; Francis Bowes Sayre, '*Mens rea*' (1932) 45(6) *Harvard Law Review* 974, 976.
[25] Theodore Frank Thomas Plucknett, *A Concise History of the Common Law* (5th ed. The Lawbook Exchange 2001), 444 – 445; citing 14 Edw. III Stat. 1 c.4 (1340).
[26] Frederick Pollock and Frederic William Maitland, *The History of English Law Before the Time of Edward I* (2nd ed. The Lawbook Exchange 2008), 45.

general laws of the land, and it was this which significantly imported notions of moral guilt and the beginning of the development of *mens rea* as an essential component of criminal liability.[27] Particularly influential were the writings of Henry Bracton, who wrote 'it is will and purpose which mark *maleficia*' and 'a crime is not committed unless the intention to injure exists.'[28]

Over following centuries, the original singular notion of *mens rea* as an evil motive was gradually transformed through attempts to 'identify specific states of mind required for the commission of particular offenses.'[29] The seminal jurist Edward Coke is credited with developing the concept of "malice aforethought" in murder to incorporate both express and implied malice. The latter included various such examples as killing the King's officers or watchmen whilst in the execution of their duties, or the killing of prisoners by their jailors. Thus, the original notion of *mens rea* as a purely mental element similarly transformed to include attributions of malice based not on the subjective mental state of the killer but on the status of the victim. This seemingly nuanced approach to *mens rea* can be found on stark display in the case of *Gregson v Gilbert*,[30] which substantially concerned an insurance claim for property lost at sea. However, the "property" in question actually consisted of more than 150 slaves taken from Africa, who had been deliberately thrown overboard when the ship came into difficulties and suffered diminished resources at sea. Whatever the moral condemnation of murder, the intentionality of the ship's captain, and the killing of other people resulting from his actions, no criminal offence had been committed owing to the status of the victims.

The eighteenth and nineteenth centuries finally saw the full development of different specific formulations of *mens rea* which would be more or less recognisable today. The shift from a single conception of the evil mind to various different formulations of subjective states of mind was further accompanied by the introduction of features such as crimes of strict liability (not requiring any *mens rea* at all), the incorporation of negligence

[27] Sayre (1932), 982 – 987.
[28] Henry Bracton, *Bracton on the Laws and Customs of England* (Thorne S. E. (trns.) Belknap Press 1968), 290.
[29] Gardner (1993), 667.
[30] *Gregson v Gilbert* (1783) 3 Doug. KB 232.

into criminal responsibility (which does not refer to any state of mind of the defendant), and a 'growing reliance on the "presumption of malice and intent".'[31] Thus, as Singer explains, by the early twentieth century the criminal law was 'no longer concerned with a general "*mens rea*", but only with a much more specific, constrained question of whether the defendant's conduct reflected the specific mental state required.'[32]

Indeed, it is clear that the concept of moral blameworthiness has continued to lose favour as the foundation of *mens rea*, whatever the historical connection. *Mens rea* has diverged away from a singular notion of malicious intent to encompass a range of different states of mind – intention, recklessness, knowledge and belief *etc.* – as well as non-mental states such as where negligence operates as *mens rea*. In so doing, as Ormerod and Laird explain, for some crimes "fault" is no longer limited to solely encompass subjective states of mind and, instead, the prosecution 'have to prove only that the defendant did not behave in a way *a reasonable person would* and, in the case of a resultant crime, thereby caused the proscribed result.'[33]

Nevertheless, it is submitted that *mens rea* too often persists in denoting moral blameworthiness, even if the courts are concerned with proof of the existence of specific subjective states of mind as opposed to any general notion of wrongdoing. For example, Simester writes that 'culpability is a particular kind of moral evaluation' and 'to blame someone is to make a *moral assessment* of that person in respect of their action.'[34] Allen and Edwards write that *mens rea* 'imports a notion of culpability or *moral blameworthiness*.'[35] In relation to US law, Hall writes that the 'relevant moral judgment implied in the penal law is absolute: no matter how good the actor's motives, since he voluntarily (*mens rea*) committed a penal harm he is to some degree morally culpable.'[36]

---

[31] Richard G. Singer, 'The resurgence of *mens rea*: I – Provocation, emotional disturbance, and the model penal code' (1986) 27(2) *Boston College Law Review* 243, 243 – 244.
[32] *Ibid*., 244.
[33] David Ormerod and Karl Laird, *Smith, Hogan, and Ormerod's Text, Cases, and Materials on Criminal Law* (13th ed. Oxford University Press 2020), 97 (emphasis added); Child and Ormerod (2019), 84 – 85.
[34] Andrew P. Simester, 'A disintegrated theory of culpability' in Baker D. J. (ed.), *The Sanctity of Life and the Criminal Law: The Legacy of Glanville Williams* (Cambridge University Press 2013), 180 (original emphasis).
[35] Michael Allen and Ian Edwards, *Criminal Law* (15th ed. Oxford University Press 2019), 76.
[36] Jerome Hall, *General Principles of Criminal Law* (2nd ed. The Lawbook Exchange 2005), 94.

And Singer writes, 'we may soon find that the criminal law has been restored to its primary, indeed its only, focus – imposing social stigma on those who knowingly, purposely, or recklessly act in disregard to social and *moral duties*.'[37]

*

Contrary to these preceding quotations connecting *mens rea* to moral blame, above, Card and Molloy correctly assert that '*mens rea* has nothing necessarily to do with notions of an evil mind, moral fault, or knowledge of the wrongfulness of the conduct.'[38] The authors offer a number of arguments in support, some of which are developed further in this section. In brief, however, they submit: that proof of the absence of moral fault *per se* does not in itself provide any legal defence;[39] and that a person's motives, whether good or bad, are irrelevant to the question of criminal responsibility.[40] Furthermore, there is no defence in ignorance or mistake as to the law (albeit this may mitigate punishment);[41] nor any defence because a defendant 'did not personally consider his conduct to be immoral or know that it was regarded as immoral by the bulk of society.'[42] This latter point in particular has been reflected in the Supreme Court's shift from a purely subjective to an objective interpretation of the *mens rea* of dishonesty, which prevents defendants from asserting in their defence that they did not realise that their actions were considered to be dishonest.[43]

Perhaps the most common criticism of linking *mens rea* with moral blame is the argument that there is a disconnection in general between the law and morality; that these two entities are not one and the same thing. Stated more precisely, whilst it is true that law

[37] Richard G. Singer, 'The resurgence of *mens rea*: III – The rise and fall of strict criminal liability' (1989) 30(2) *Boston College Law Review* 337, 408 (emphasis added).
[38] Richard Card and Jill Molloy, *Card, Cross & Jones Criminal Law* (22nd ed. Oxford University Press 2016), 77.
[39] *R v Yip Chiu-Cheung* [1995] 1 AC 111, 117 – 118; *R v Kingston* [1995] 2 AC 355, 364 – 366; *R v Dodman* [1998] 2 Cr App R 338.
[40] *Chandler v Director of Public Prosecution* [1964] AC 763; *Hills v Ellis* [1983] QB 680; *Attorney-General's Reference (No. 1 of 2002)* [2002] EWCA Crim 2392; *Attorney-General v Scotcher* [2005] UKHL 36.
[41] *Johnson v Youden* [1951] 1 KB 544; *Churchill v Walton* [1967] 2 AC 224, 226; *Paul v Ministry of Posts and Telecommunications* [1973] RTR 245.
[42] Card and Molloy (2016), 116.
[43] See *Ivey v Genting Casinos (UK) Ltd. t/a Crockfords* [2017] UKSC 67.

and morality may at times govern similar matters or respond to similar questions, it is submitted that there is no *necessary* connection between the two. On the one hand, there exist actions and activities that many people would consider morally objectionable but yet the law fails to ostensibly criminalise, such examples including certain product testing on animals (only regulated for particular species and procedures),[44] racial, sexist and other forms of discrimination (only regulated in civil law),[45] and tax avoidance (only prosecuted when crossing a threshold into criminal tax evasion).[46] On the other hand, there exist numerous criminal offences that many other people would not consider to be morally objectionable at all; or, even, consider that the criminalisation of such offences is itself a moral wrong. Such examples include the criminalisation of homelessness (termed as "vagrancy"),[47] of the possession of certain drugs,[48] and of certain end of life practices including assisted suicide and euthanasia.[49]

If the law criminalises moral wrongs, therefore, which or whose morality is being imposed? Certainly, in the United Kingdom at least, the criminal law does not impose a system of religious canon law upon the general public (even if particular historical influences such as Sunday trading laws do remain). Indeed, the law of the UK recognises and protects the religious pluralism of British society, and does not punish such religious offences as adultery or blasphemy. One common proposal is that the law criminalises actions according to the harm principle, *i.e.,* those actions which cause harm to others or society in general.[50] This principle is largely attributable to the seminal philosopher John Stewart Mill who wrote:

> 'The only purpose for which power can be rightfully exercised over any member of a civilised community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant.

[44] Animals (Scientific Procedures) Act 1986.
[45] Equality Act 2010.
[46] Criminal Finances Act 2017.
[47] Vagrancy Act 1824.
[48] Misuse of Drugs Act 1971; Psychoactive Substances Act 2016.
[49] Suicide Act 1961; *R v Cox* (1992) 12 BMLR 38.
[50] Jonathan Herring, *Criminal Law: Text, Cases, and Materials* (9th ed. Oxford University Press 2020), 18 – 23.

> He cannot rightfully be compelled to do or forbear because… in the opinion of others, to do so would be wise, or even right.'[51]

A similar criticism may again be levied against the harm principle as an underlying morality for the criminal law. There are actions and activities that are criminalised without any clear link to public harm, such as the criminalisation of fox hunting[52] and, again in this case, the criminalisation of homelessness and of any animal testing for product safety. Lord Devlin has proposed that, rather than necessarily relating to physical harm, there is a 'moral cement' which assists in keeping a society together, and that the 'extent of disgust felt by society at a particular kind of activity would indicate whether it challenged a fundamental value that underpinned society.'[53] However, again, where is such moral cement to be found in a multi-cultural and multi-faith society such as the UK; and, moreover, it may be questioned whether emotional disgust is even a worthy foundation for moral value. As Herring notes, 'many people experience great disgust at the picking of a nose, but that does not indicate that it reflects a fundamental moral principle.'[54]

Arguably the strongest claim that might be made is that the criminal law reflects the broader morality of a nation's people as implemented through the process of representative democracy. However, again, there are a number of challenges to this view, in particular as it relates to common law jurisdictions. First, instrumental changes to the criminal law can clearly be made by judges in the courts. Two prominent examples include the historical exclusion of the defence of necessity for the offence of murder,[55] and the relatively modern decision (and reversal of the pre-exiting position) that rape can occur within a marriage.[56] The latter example in particular reveals why the codification of the criminal law would not necessarily resolve this first criticism; the offence of rape

[51] John Stewart Mill, *'On Liberty' and Other Writings* (Collini S. (ed.) Cambridge University Press 1989), 13.
[52] Hunting Act 2004.
[53] Herring (2020), 21; citing Patrick Devlin, *The Enforcement of Morals* (Liberty Fund 2010); see also Steven Wall, 'Enforcing morality' (2013) 7(3) *Criminal Law and Philosophy* 455.
[54] Herring (2020), 21.
[55] *R v Dudley and Stephens* (1884) 14 QBD 273.
[56] *R v R* [1991] UKHL 12.

already existed, yet the courts still considerably changed the circumstances in which the offence could be committed.[57]

Second, and related to the first objection, there exist a number of offences which still today do not possess a statutory basis, the most surprising being the offence of murder which remains largely based upon jurisprudence from the 17th century jurist Edward Coke.[58] Whilst it is generally posited by the courts themselves that it is no longer appropriate for the judiciary to create new common law criminal offences,[59] those which do exist, (along with the court's significant power to amend the interpretation and application of statutory offences), nonetheless continue to disrupt any link between the criminal law reflecting public morality as enacted through democratic processes.

There are a number of instances where evidence suggests that the views and opinions of the public are not being properly implemented by their representatives in Parliament. The most prominent example relates to end-of-life practices such as assisted suicide and euthanasia. One of the largest public polls conducted on the subject in the UK in 2015 revealed that as much as 82% of the population supported some form of legal reform of the blanket prohibition against end-of-life assistance, including surprisingly high support from amongst religious groups that have traditionally opposed end-of-life assistance in favour of the sanctity of life.[60] Notwithstanding such public opinion, successive attempts to reform the law have been brought before Parliament and repeatedly voted down in spite of public favour for reform.[61] Meanwhile, several challenges to the existing law have been brought before the courts, and a number of justices of the Supreme Court have commented their view that the current blanket prohibition against assisted suicide and euthanasia may be in breach of fundamental human rights, even threatening Parliament

[57] See further Marianne Giles, 'Judicial law-making in the criminal courts: the case of marital rape' (1992) (Jun) *Criminal Law Review* 407.
[58] Tony Storey, *Unlocking Criminal Law* (7th ed. Routledge 2020), 7 – 8.
[59] *R v Price* (1884) 12 QBD 247; *R v Coney* (1882) 8 QBD 534, 550; see generally A. T. H. Smith, 'Judicial law making in the criminal law' (1984) 100(1) *Law Quarterly Review* 46.
[60] Populus, *Dignity in Dying Poll* (2015) [Online] <http://www.populus.co.uk/wp-content/uploads/2015/12/DIGNITY-IN-DYING-Populus-poll-March-2015-data-tables-with-full-party-crossbreaks.compressed.pdf> accessed 20th October 2020.
[61] Recent examples include the Patient (Assisted Dying) Bill 2002; Terminally Ill Bill 2004; Assisted Dying Bill 2014; Assisted Dying Bill (No. 2) 2015-16; Assisted Dying Bill 2016-17; Assisted Dying Bill 2021.

with a declaration of incompatibility unless the issue is reconsidered.[62] Clearly, therefore, public morality is not always reflected in the actions that Parliament chooses to criminalise or not.

On this latter point, it might be argued that a conceptual separation between law and morality exists for moral reasons, thereby reintroducing a necessary connection between the law and morality. For example, it would generally be regarded as highly immoral for a capable bystander to do nothing to assist another from grave danger – *e.g.,* a child drowning in a shallow pond – when to do so would pose little danger to that bystander themselves. Yet, in UK law, no such legal duty arises to assist endangered strangers in either civil or criminal law,[63] albeit such a duty may arise within special identified relationships, such as with regards to parents and teachers towards children in their care, the occupiers of property in relation to lawful visitors, or where a stranger has voluntarily assumed such responsibility towards another, *etc*.[64]

The House of Lords in *Stovin v Wise* offer a number of arguments in support of this position: the "political" (and also arguably somewhat moral) argument that it is a lesser invasion of personal liberty to require in law that people take care not to injure others by their actions, as opposed to a more invasive positive duty to rescue or protect others; the "moral" question of why any particular individual would be chosen for prosecution for not rendering assistance when any number of a large and indeterminate class of people might have been available to help; and the "economic" (and, again, arguably somewhat moral) argument that activities should bear their own costs, such that individuals bear the cost of their own danger as opposed to liability being shifted onto "innocent" bystanders.[65]

Thus, it might be argued, even though the law *permits* the ostensibly immoral position of the "callous" bystander who refuses to assist the child drowning in a shallow pond, the reasons for the absence of so-called "good Samaritan" laws in the UK are arguably moral

[62] *R (on the application of Nicklinson) v Ministry of Justice* [2014] UKSC 38.
[63] *Stovin v Wise* [1996] AC 923; *Yuen Kun Yeu v Attorney-General of Hong Kong* [1988] AC 175.
[64] See further *Smith v Littlewoods* [1987] UKHL 18.
[65] *Stovin* [1996], 943 – 944.

– *i.e.,* the law takes the position that it is immoral to impose good Samaritan laws, reintroducing a degree of necessary connection between law and morality. The problem with this view, however, is the ready existence of good Samaritan laws in jurisdictions around the world, most notably across continental Europe such as in France, Belgium and Germany, where the respective criminal codes make it a criminal offence for the callous bystander to refuse to assist another in danger when to do so would not disproportionately endanger themselves.

Presumably in such jurisdictions, the moral good of imposing duties to rescue others in danger is deemed to exceed such moral counterarguments as are presented to underlie the UK position in *Stovin v Wise*.[66] In any event, the example again undermines the argument that there exists a *necessary* connection between law and morality, whether one is considering those things that the law *does* select to regulate, or those things that are to be deemed for moral reasons to be outside of the purview of the law. In either case, the claim of any necessary connection between law and morality – and, in particular, that *mens rea* operates to denote moral blame – invites an inescapable moral relativism when considered against the vast differences in criminal prohibitions between different cultures, societies and histories.

As a final point on the matter, it is submitted that morality and moral blame provide a generally incoherent foundation for the criminalisation of different actions and behaviour. In particular, it is difficult to reconcile why a given action might be regarded as immoral one day and not the next. Phrased differently, was rape within a marriage any more morally defensible in 1990 prior to the landmark ruling of the House of Lords in *R v R* in 1991? Equally, was homosexuality any more morally reprehensible in 1966 prior to the decriminalisation of sexual acts between two men under the Sexual Offences Act in 1967? And, if murder is criminalised because it is arguably the most heinous offence, what explains the disparate and varied application of the offence of murder over the course of

[66] See further Martin Vranken, 'Duty to rescue in civil law and common law: Les extremes se touchent?' (1998) 47(4) *International and Comparative Law Quarterly* 934; Jan M. Smits, 'The good Samaritan in European private law: On the perils of principles without a programme and a programme for the future' (Inaugural lecture, Maastricht University, 19 May 2000) <https://core.ac.uk/download/pdf/231273801.pdf> accessed 16 October 2022.

centuries, from a civil wrong to a criminal offence, and a criminal offence which goes from excluding to including various categories of victim (including entire races during the slave trade). It is submitted that moral blame as a foundation for criminalisation can only be justified with a species of moral relativism that most people in society would struggle to ascribe to or accept. Moral tenets often provide (near) absolute precepts of behaviour – do not kill, do not steal, *etc.* In contrast, the law recognises considerably greater flexibility, both with regards to the specific actions that are or are not subject to criminalisation over the course of time, and with regards to the way that criminal offences are each interpreted and applied in different circumstances.

For example, not only is it clear from the discussion at the introduction of this section, above, that the offence of murder has changed considerably over centuries, but even today there remain certain exceptions to what would at first appear to be the most absolute of legal and moral proscriptions. Most obviously, killing another can indeed be justified in law in situations of self-defence. Although the defence requires that only reasonable force is used by the defendant, such force as that which kills an assailant may indeed be regarded as reasonable in certain life-or-death scenarios.[67] The comparatively modern case of *Re A (conjoined twins)*[68] illustrates the broader point, in which a pair of conjoined twins were both certain to die without an operation to surgically separate them, but the necessary operation would inevitably result in the death of one twin whilst saving the other.

On the one hand, it is well established that the defence of necessity would be unavailable to the charge of murder,[69] the technical elements of which would almost certainly be established in relation to the twin destined to die. On the other hand, it was clear that the applicable legal test – acting in the best interests of each child – was in conflict depending upon which twin's interests were considered. Ultimately, the Court of Appeal did fall back upon an argument of necessity;[70] however, they restricted their finding to the very

[67] *R v Chisam* (1963) 47 Cr App R 130, 133; *R v Rose* (1884) 15 Cox 540.
[68] *Re A (conjoined twins)* [2001] Fam 147.
[69] *Dudley and Stephens* (1884).
[70] Birju Kotecha, 'Necessity as a defence to murder: an Anglo-Canadian perspective' (2014) 78(4) *Journal of Criminal Law* 341, 343 – 345.

specific facts of the case at hand and, furthermore, avoided making reference to the legal defence of necessity *per se*, instead referring to the 'lesser of two evils.'[71] This clearly displays the degree of flexibility required in law, which is not necessarily available when considering a more general absolutism that is associated with moral tenets and dictates.

### *8.3.1. The Importance of Denouncing Moral Blame*

Why is it significant to finally break this link between *mens rea* and moral blame? As Gordon asserts, 'moral blameworthiness is a function of the state of mind (or will) of the agent himself.'[72] Thus, whereas the courts are no longer concerned with any singular concept of moral wrongdoing, this historical foundation of *mens rea* in moral blameworthiness continues to be reflected in the different formulations of *mens rea* as subjective states of mind. That is to say, the law imposes liability for criminal acts committed in a particular subjective state of mind because it is this which reflects the fact that a defendant was morally blameworthy, as the defendant should have chosen to act differently (applying the assumption of conscious control). And yet, as this chapter of the thesis has explored, a number of assertions in the previous statement are called into question, specifically the assumption of conscious online control over decisions and their resultant actions, and the various issues identified with relying upon subjective mental states as a foundation for establishing legal responsibility.

There are two final critiques that may be raised against moral blame underlying *mens rea*; namely, the argument that moral blame denotes a metaphysical ability to have chosen to act differently within the same set of circumstances (summarised as "*ought implies can*"), and the argument that moral blame justifies or even requires retributive punishment. Starting with the first argument, it has previously been stated that the legal presumption of volition relies upon the unproven assumption that volition denotes a degree of online conscious control over thought and action. The argument then follows that, because people possess this assumed ability for conscious control then, in any given situation where they decide to commit a criminal action, they are responsible because they ought

[71] *Re A (conjoined twins)* [2001], 203 & 239.
[72] Gerald H. Gordon, 'Subjective and objective *mens rea*' (1974) 17 *Criminal Law Review* 355, 355.

to have chosen differently. Thus, *vice versa*, moral proscriptions that a person ought to have acted differently in a given situation implies that they could indeed have done so.

The ethical formula of *ought implies can* is attributed to Immanuel Kant, who writes 'if the moral law commands that we *ought* to be better human beings now, it inescapably follows that we must be *capable* of being better human beings,'[73] and 'the action to which the "*ought*" applies must indeed be possible under natural conditions.'[74] As this relates to moral blame and conscious control of action, the problem is that if criminal offences reflect moral wrongs and *mens rea* reflects moral blameworthiness, a moral/legal command not to commit a certain act implies that the same individual placed twice within the same set of circumstances has an ability to choose to act differently on each occasion (*i.e.,* the principle of alternative possibilities).[75] Conversely, the deterministic worldview that is assumed at the outset of this thesis posits that a person acting one way within a given set of circumstances (*i.e.,* encapsulating all the *causes* of their present behaviour) would necessarily act the same if time was reversed and they were placed back in exactly the same set of circumstances. As Kahn explains in a number of steps:

> '[T]he argument begins with the premise the (1) an agent is blameworthy for performing a given act A only if s/he has an obligation not to perform A. But (2) if *ought implies can*, then an agent has an obligation not to perform A only if s/he is able not to perform A. It follows immediately that (3) if *ought implies can*, then an agent is blameworthy for performing a given act A only if s/he is able not to perform A. But (4) if an agent is able not to perform A, then s/he is able to do otherwise (than A), whence it

[73] Immanuel Kant, *Religion within the Boundaries of Mere Reason: And Other Writings* (Wood A. and Giovanni G. (eds.) Cambridge University Pres 2018), 81 (original emphasis).
[74] Immanuel Kant, *Critique of Pure Reason* (2nd ed., Smith N. K. (trns.) Palgrave Macmillan 2007), 473.
[75] See further David Copp, '"Ought" implies "can" and the derivation of the principle of alternate possibilities' (2008) 68(1) *Analysis* 67; David Copp, '"Ought" implies "can", blameworthiness, and the principle of alternate possibilities' in Widerker D. and McKena M. (eds.), *Moral Responsibility and Alternative Possibilities* (Ashgate Publishing 2003).

follows that (5) if *ought implies can*, then an agent is blameworthy for performing a given act A only if s/he is able to do otherwise.'[76]

This challenge can be avoided if moral blame is abandoned as the basis of subjective *mens rea* and criminal responsibility; and, indeed, there are many reasons to suggest that moral commands and criminal proscriptions are two very different things. Most obviously, whereas moral obligations do not possess any formal means of enforcement, attached to criminal offences are their associated sentences backed by the authority of law. That is to say, whereas there may be no formal repercussions for the breach of a moral obligation, committing a criminal offence results in some form of prescribed punishment which, if not complied with, may be compelled by the State by force (*e.g.,* through enforced incarceration). In this regard, the criminal law does not strictly tell people what they ought to do (thus avoiding the issue of *ought implies can*); rather, it states what actions are prohibited and, more importantly, warns of the response of the State if and when such offences are committed. Moral prohibitions are "merely" normative in prescribing how people *ought* to act, whereas legal prohibitions are coercive in prescribing how people *must* act in order to avoid the threat of punishment.

Is this distinction between moral obligations denoting "*ought*" and criminal laws denoting "*must*" meaningful or artificial? One common objection against overly harsh criminal sentences and capital punishment in particular is that the typical person is rarely weighing up the deterrent effect of potential sentences *in the moments of actually committing a criminal offence*. Rather, so far as they may be concerned with the law at all, more often that concern is about evading capture for the offence they are in the process of committing, rather than reconsidering their actions during the moment of commission because they have suddenly remembered that it is a criminal act carrying the threat of punishment. This is not to suggest that the law does not act as a meaningful deterrent, but that its impact upon the behaviour of people engaged in criminal activities rarely occurs at the moment of committing an offence, by which stage the law's deterrent impact has either been

[76] Samuel Kahn, *Kant, Ought Implies Can, the Principle of Alternative Possibilities, and Happiness* (Lexington Books 2019), 149 – 150; citing David Widerker, 'Frankfurt on "ought implies can" and alternative possibilities' (1991) 51(4) *Analysis* 222.

effective (and the individual does not initiate the offence) or it has not (and the offence is committed). Instead, the deterrent effect of punishment which attaches to the law's *obligatory* proscriptions (but does not so attach to moral statements about what a person *ought* to do) either takes effect before the commission of an offence during the processes of deciding whether or not to commit a particular criminal (*i.e.,* general deterrence), or after the fact when imposed punishment dissuades the offender from repeating their criminal actions in the future (*i.e.,* special deterrence).

A number of studies may be read to indirectly support the aforementioned arguments. Firstly, 'no credible and consistent body of evidence has been found to support the conclusion that harsher sentences… achieve marginal deterrent effects on crime.'[77] These conclusions are drawn from a review of more than 15 studies conducted since 1975 investigating the impact of harsher sentencing. From some of the individual studies considered, specific comments assert that the evidence for the deterrent effect of higher sentencing 'falls well short of being a theory that should continue to enjoy the allegiance of criminologists,'[78] and 'it is time to accept the null hypothesis [that] variation in the severity of sanctions is unrelated to levels of crime.'[79] Secondly, further research has investigated the little-studied question of why those people who commit serious crimes nevertheless tend to comply with the law the majority of the time.[80] Collecting survey data from 141 serious gun-crime offenders in the US, Papachristos, Meares and Fagan conclude that offenders are 'more likely to comply with the law when they believe (a) in

[77] Cheryl Marie Webster and Anthony N. Doob, 'Searching for sasquatch: Deterrence of crime through sentence severity' in Petersilia J. and Reitz K. R. (eds.), *The Oxford Handbook of Sentencing and Corrections* (Oxford University Press 2012), 174; citing Andreas von Hirsch, Anthony E. Bottoms, Elizabeth Burney and P. O. Wikstrom, *Crime Deterrence and Sentencing Severity* (Hart Publishing 1999); Shawn Bushway and Michael A. Stoll (eds.), *Do Prisons Make Us Safer: The Benefits and Costs of the Prison Boom* (Russell Sage Foundation 2009); Robert Apel and Daniel S. Nagin, 'General deterrence' in Tonry M. H. (ed.), *The Oxford Handbook of Crime and Criminal Justice* (Oxford University Press 2011).

[78] Travis C. Pratt, Francis T. Cullen, Kristie R. Blevins, Leah E. Daigle and Tamara D. Madensen, 'The empirical status of deterrence theory: A meta-analysis' in Cullen F. T., Wright J. P. and Blevins K. R. (eds.), *Taking Stock: The Status of Criminological Theory: Volume 15* (Routledge 2017), 385.

[79] Anthony N. Doob and Cheryl Marie Webster, 'Sentence severity and crime: Accepting the null hypothesis' (2003) 30(1) *Crime and Justice* 143, 143.

[80] Andrew V. Papachristos, Tracy L. Meares and Jeffrey Fagan, 'Why do criminals obey the law? The influence of legitimacy and social networks on active gun offenders' (2012) 102(2) *Journal of Criminal Law and Criminology* 397.

the legitimacy of legal actors, but especially the police, and (b) that the substance of the law is consistent with their own moral schedules' or values.[81]

The findings by Papachristos, Meares and Fagan suggest that the deterrent effect of the law is concerned with the broader legitimacy of the law and its institutions (such as the police) and, crucially, the values which they reflect. These factors, in turn, impact upon people's attitudes towards breaking the law. Thus, it may again be appreciated that the deterrent impact of the law is not taking place at the moment of criminal activity, but is more closely associated with views, beliefs and attitudes towards the law formed over a longer period of time – before (*i.e.,* general deterrence) or after (*i.e.,* special deterrence) an offence has actually been committed – which themselves influence the likelihood of an individual committing a (further) criminal offence. In this regard, Tonry notes that most crimes are typically not highly calculated but, rather, are more likely to be impulsive and committed 'under the influence of drugs, alcohol, peer influences, powerful emotions, or situational pressures.' [82] Such influences are reasoned to negate much of the opportunity that might otherwise have been available for the law to exert a deterrent effect *in the moment* of breaking the law and committing an offence.

A final argument supporting the meaningful distinction between moral obligations denoting "*ought*" and criminal laws denoting "*must*" lies in the implications that follow this distinction with regards to the question of punishment. Specifically, the concept of moral blameworthiness proceeds to support retributive theories of punishment which are significantly more difficult to justify in an entirely deterministic worldview, within which an individual placed twice in precisely the same set of circumstances (*i.e.,* causes) would be bound to act in the same way. The retributive argument is fully expressed in the work of Michael S. Moore,[83] for whom the 'criminal law is a functional kind whose function is to attain retributive justice' which 'demands that those who deserve punishment get it.'[84]

---

[81] *Ibid*., 400.

[82] Michael H. Tonry, *Crime and Justice: A Review of Research, Volume 37* (University of Chicago Press 2008), 2.

[83] See Michael S. Moore, 'Four reflections on law and morality' (2007) 48(5) *William & Mary Law Review* 1523; Michael S. Moore, 'A tale of two theories' (2009) 28(1) *Criminal Justice Ethics* 27; Michael S. Moore, 'The various relations between law and morality in contemporary philosophy' (2012) 25(4) *Ratio Juris* 435.

[84] Michael S. Moore, 'Liberty's constraints on what should be made criminal' in Duff R. A., Farmer L., Marshall S. E., Renzo M. and Tadros V. (eds.), *Criminalization: The Political Morality of the Criminal Law*

Therefore, he argues, the criminal law must 'punish all and only those who are morally culpable in the doing of some morally wrongful action.'[85]

Moore's retributive theory is founded upon the principle of moral wrongdoing, for which he relies upon the principle that choices are volitional and, consequently, controlled.[86] In particular, Moore emphasises the capacity for conscious control over decisions and actions as central to his notion of volition, which extends from control over our broad intentions right down to the fine motor movements that bring those intentions into existence through action. As he writes, 'the conscious experience of acting… is one of the crucial experiences needed to be a person at all' and, 'even when [a] dim awareness of the movement is absent, it is none the less accessible to consciousness.'[87]

In light of the conclusions from Part One of this thesis and the preceding discussion in this chapter, a number of problems with such a retributive theory as Moore's ought to become plain. First, Moore's concept of volition (which underpins his broader theory of retributivism) is rooted in *conscious* control over thoughts and actions or, at the very least, conscious accessibility to the same. However, as has been discussed in this thesis, it is perfectly possible (if not highly likely) for goals and intentions to be instigated, processed and put into action entirely outside of conscious awareness. Equally, it is likely that both the decision of *what* to do and a final decision of *whether* or not to veto an action are each the product of unconscious activity in the brain prior to entering conscious awareness (if they so enter conscious awareness at all), thus precluding consciousness *per se* as the source of self-control over our behaviour.

It is also submitted that Moore overestimates the involvement of consciousness in the control of fine motor movements. He uses the example of somebody throwing a ball to hit a target, suggesting that fine control over motor actions is at least accessible to

(Oxford University Press 2014), 191; see also Michael S. Moore, *Placing Blame: A General Theory of the Criminal Law* (Oxford University Press 1997), Chs. 2 – 4.

[85] Moore (1997), 33 & 35.

[86] Michael S. Moore, *Act and Crime: The Philosophy of Action and Its Implications for Criminal Law* (Oxford University Press 2010), 151.

[87] *Ibid*., 154; see also Harry Kalven, 'Insanity and the criminal law – A critique of *Durham v United States*' (1955) 22(2) *University of Chicago Law Review* 317.

consciousness. When aiming at a target, the brain must indeed calculate a precise trajectory and then match fine motor movements in order to throw the ball upon this trajectory with the correct force and direction *etc*. However, few people would be able to consciously access and describe the actual angles that such a trajectory requires, or the actual degree of force that needs to be applied to the ball, or the exact direction in degrees that the ball needs to be thrown. Whilst we may indeed be consciously aware of engaging a throwing motion in order to hit a particular target, the actual calculations and fine motor control exerted by the brain in order to achieve this goal are the product of decades of life-experience in controlling motor movements and matching such movements with plans which pursue goals.

Control over motor actions is something that is gained through years of practise, just as a musician must practise extensively in order to perform the fine movements necessary to play an instrument proficiently. Such control may reasonably be enhanced through conscious concentration – again, such as when a musician first learns a new piece of music. Here, consciousness may be reasoned to improve control by providing greater time and mental resources dedicated to the mechanisms in the brain that exert self-control over actions. Crucially, however, those mechanisms continue to exist and function whether or not they are subject to being exercised consciously; otherwise, we would lose control over our actions the moment we ceased giving them conscious attention.

Second, Moore's claims that the criminal law ought to 'punish all and only those who are morally culpable in the doing of some morally wrongful action' draws the link between the criminal law and moral wrongdoing. However, as discussed above in this chapter, there is no necessary link between actions that the criminal law prohibits and the inherent morality or immorality of those actions. The law both criminalises certain activities that can scarcely be considered to be immoral and, equally, fails to criminalise other actions that patently are immoral. Moral blame as an underlying justification for criminalisation struggles to provide a coherent account for why the law (and, therefore, presumably morality and moral value) is so changeable, and ultimately can only be supported with a form of moral relativism which would be objectionable to a great many people. Fundamentally, if morality is the underlying basis for the criminal law, there is no entirely

satisfactory answer to the question of which moral code is to be implemented in law, with even appeals to the democratic process providing an inadequate answer to this question.

Third and finally, it is submitted more generally that retributive theories of criminal law such as Moore's do not hold up within the deterministic worldview assumed at the outset of this thesis and supported by the evidence considered throughout. That is to say, retributivism requires that people are able to choose to act otherwise than they actually do given the exact same set of circumstances (*i.e.,* causes) – this is a subtle variation on what is otherwise known as the principle of alternate possibilities, and is discussed more fully in chapter thirteen of this thesis, below.[88] Conversely, however, determinism requires that the exact same set of causes would result in the same effects, which applies as equally to human behaviour as it does to chemical reactions or the motion of planets. As Kelly writes, retributivism 'is the view that justice requires the punishment of criminal wrongdoers apart from the (further) social benefits a system of punishment might bring. The case for this notion of justice is built on reactive attitudes that presuppose a wrongdoer's moral capacity to have acted as morality demands. If we drop the assumption that offenders always have this capacity, we must re-evaluate the aims of punishment.'[89]

The challenge to retributivism posed by determinism (or arguments otherwise described as "free will scepticism") has been widely explicated.[90] Taggart describes succinctly,

> '(i) a necessary condition for an actor's desert is that she conduct herself in some way and be morally responsible for that conduct; (ii) a necessary

[88] One of the most influential writers on the issue of the principle of alternate possibilities – Harry Frankfurt – strongly asserts that this principle is not in fact a necessary requirement for moral responsibility. This position is more fully engaged in chapter thirteen of this thesis, below; however, for the present discussion, it is submitted that the principle is indeed a standard prerequisite for retributivism.
[89] Erin I. Kelly, 'Criminal justice without retribution' (2009) 106(8) *Journal of Philosophy* 440, 446.
[90] For example, see Joshua Greene and Jonathan Cohen, 'For the law, neuroscience changes nothing and everything' (2004) 359(1451) *Philosophical Transactions of the Royal Society: Biological Sciences* 1775; Derk Pereboom, 'Determinism al dente' (1995) 29(1) *Noûs* 21; Derk Pereboom, *Living Without Free Will* (Cambridge University Press 2001), 159 – 161; Derk Pereboom, 'Free will skepticism and criminal punishment' in Nadelhoffer T. A. (ed.), *The Future of Punishment* (Oxford University Press 2013); Christopher P. Taggart, 'Retributivism, agency, and the voluntary act requirement' (2016) 36(3) *Pace Law Review* 645; Christopher P. Taggart, 'Retributivism, ultimate responsibility, and agent causalism' (2019) 54(3) *Tulsa Law Review* 441; Bruce Waller, *Against Moral Responsibility* (Massachusetts Institute of Technology Press 2011).

> condition for an actor to be morally responsible for her conduct is that she have the right sort of control over her conduct; (iii) a necessary condition for an actor to have the right sort of control over her conduct is that agent causalism be true; therefore, (iv) a necessary condition for an actor's desert is that agent causalism be true.'[91]

Agent causalism regards the individual agent acting voluntarily as the originating cause of their free, morally responsible actions, distinguishing them from surrounding events and circumstances, including discrete activity or states of affairs within the brain. However, this view is significantly called into question by the evidence considered throughout this thesis. More precisely, it is submitted that the decisions and actions that people exhibit are inseparable from the underlying activities in the brain and wider nervous system which ultimately produce or cause those decisions and actions, and are themselves bound to a preceding chain of deterministic causation. It follows that the foundational basis of retributive punishment is lacking which, as will be explored further in chapter thirteen of this thesis, below, can have profound consequences for the operation of the criminal justice system and theories of punishment in particular.

[91] Taggart (2016), 652.

# 9. Reconstructing *Mens Rea*

> 'Responsibility might depend on the reason that triggered a neural process culminating in action, and on whether a final check should have stopped the action. Interestingly, both decisions have a strong normative element: although a person's brain decides the actions that they carry out, culture and education teach people what are acceptable reasons for action, what are not, and when a final predictive check should recommend withholding action. A neuroscientific approach to responsibility may depend not only on the neural processes that underlie volition, but also on the brain systems that give an individual the general cognitive ability to understand how society constrains volition, and how to adapt appropriately to those constraints. A basic level of functioning of the social brain, as well as the cognitive-motor brain, is essential for our conventional concept of responsibility for action.'
>
> - Patrick Haggard, 2008.[1]

## *9.1. The Reasonableness Principle*

Following immediately from the preceding discussion, where it is argued that the link between *mens rea* and moral blame ought to be abrogated for good, it is submitted that the resulting lacuna may be replaced with the concept of "reasonableness". Put differently, it is submitted that rather than *mens rea* reflecting the moral blameworthiness of an action committed within a certain state of mind, instead *mens rea* reflects the unreasonableness of that action according to the norms of society. This principle, as an underlying foundation for *mens rea,* draws significantly from the seminal work of H. L. A. Hart, who considers that the immediate aim of criminal law is to 'announce to society that these actions are not to be done and to secure that fewer of them are done;' and that the criminal law 'sets up, in its rules, standards of behaviour to encourage certain types of conduct and

[1] Patrick Haggard, 'Human volition: towards a neuroscience of will' (2008) 9(12) *Nature Reviews Neuroscience* 934, 944.

discourage others.'[2] As Wasserstrom explains, Hart considers the aim of criminal legislation to be the 'denunciation of certain types of conduct as conduct that is not to be done.'[3] In contrast, punishment is generally justified as the means of enforcing that aim; 'the justification of punishment is that it helps to assure the general conformity to the prohibitions and requirements of the criminal law.'[4]

Why should the principle of reasonableness be preferred to that of moral blame as an underlying basis for *mens rea*? Three key arguments may be offered which broadly reflect the objections raised against morality as the underlying standard in section 8.3, above. First, it was submitted that there is a fundamental disconnect between what the criminal law proscribes and what is subject to moral judgment; that is to say, the law sometimes criminalises actions that are otherwise considered to be moral by some and, equally, sometimes fails to criminalise other actions that are widely considered to be immoral by others. The law is not in the business of prescribing any particular moral doctrine, nor are the courts concerned specifically with adjudicating the morality or immortality of conduct. This latter point in particular has been stressed in jurisprudence; for example, Lord Bingham states that the courts have the 'duty of resolving issues of law properly brought before it' but are not 'entitled or fitted to act as a moral or ethical arbiter.'[5] And, writing extrajudicially, Sir James Munby – then President of the Family Division of UK courts – asserts that 'the days are past when the business of the judges was the enforcement of morals or religious belief.'[6]

This is not to say that judges never determine moral issues; naturally, there can and often will be crossover between legal and moral questions. Nonetheless, in such circumstances the judiciary remain cautious to highlight that neither they nor the courts are *moral* arbiters – they are concerned with the law. For example, in *Airedale NHS Trust v Bland,*[7]

[2] H. L. A. Hart, 'Prolegomenon to the principles of punishment' in Hart H. L. A. (ed.), *Punishment and Responsibility: Essays in the Philosophy of Law* (2nd ed. Oxford University Press 2008), 6.
[3] Richard A. Wasserstrom, 'H. L. A. Hart and the doctrines of *mens rea* and criminal responsibility' (1967) 35(1) *University of Chicago Law Review* 92, 107.
[4] *Ibid*., 107 – 108.
[5] *R (on the application of Pretty) v Director of Public Prosecutions* [2002] 1 AC 800, 809.
[6] James Munby, 'Law, morality and religion in the family courts' (2014) 16(2) *Ecclesiastical Law Journal* 131, 133.
[7] *Airedale NHS Trust v Bland* [1993] AC 789.

which concerned the inherently moral question of distinguishing mercy killing from the discontinuation of life sustaining treatment, Lord Brown-Wilkinson commented 'it is not for the judges to seek to develop new, all embracing principles of law in a way which reflects the individual judges' moral stance when society as a whole is substantially divided on the relevant moral issues.'[8] Similarly, in the same case, Lord Mustill commented that 'adversarial proceedings, even with the help of an *amicus curiae*, are not the right vehicle for the discussion of this broad and highly contentious moral issue, nor do I believe that the judges are best fitted to carry it out.'[9] So far as the law does regulate questions of morality, it is generally recognised that Parliament is the correct forum to determine those immoral actions that may become subject to the law, whereas the courts are principally concerned with applying the law as prescribed by Parliament. As Lord Bingham writes, 'the democratic process is liable to be subverted if, on a question of moral and political judgment, opponents of the Act achieve through the court what they could not achieve in Parliament.'[10]

In contrast, the law is inherently more familiar with questions of reasonable and unreasonable conduct. Indeed, the concept of reasonableness is already endemic throughout most, if not all, areas of the law, rendering this a concept with which the courts and judiciary are eminently accustomed. For example, the standard of the *reasonable* 'man in the street' or 'man on the Clapham omnibus' is the central standard of care required under the most commonly pleaded tort of negligence.[11] In employment law, unfair dismissal from employment is determined according to the test of whether a decision to dismiss 'fell within the band of *reasonable* conduct which a *reasonable* employer could adopt.'[12] The directors of a company are under a legal duty to 'exercise *reasonable* care, skill and diligence'[13] to the standard of a '*reasonably* competent director carrying out those functions in that company.'[14] The test for negligence applicable

[8] *Ibid*., 880.
[9] *Ibid*., 890.
[10] *R (on the application of Countryside Alliance) v Attorney-General* [2008] AC 719, [45]; see also *R (on the application of Nicklinson) v Ministry of Justice* [2014] UKSC 38, [230] – [233].
[11] *Hall v Brooklands Auto Racing Club* [1933] 1 KB 205, 224.
[12] *Iceland Frozen Foods Ltd. v Jones* [1983] ICR 17, 21.
[13] Companies Act 2006, s. 174(1);
[14] Brenda Hannigan, *Company Law* (5th ed. Oxford University Press 2018), 251; citing *Re Continental Assurance Co of London plc* [2007] 2 BCLC 287.

specifically to doctors provides that, 'so long as there is a competent school of thought that supports the belief that the defendant's actions were *reasonable*, the judge will find the defendant not to have been negligent.'[15]

Equally, the standard of reasonableness can be found already in several areas of the criminal law. For example, the police must possess 'reasonable grounds' for suspicion in order to search people or vehicles,[16] to enter property without a warrant,[17] or to make an arrest without a warrant;[18] whilst conviction is dependent upon the prosecution proving their case beyond *reasonable* doubt. The defence of duress rests, in part, upon whether certain threats would have a similar effect on a 'sober person of *reasonable* firmness,'[19] and self-defence permits a defendant to act only with *reasonable* force.[20] Plainly, where the courts today generally denounce that they are moral arbiters, the concept of reasonableness is firmly entrenched in both the common law and legislation of the UK.

The second argument in favour of reasonableness over moral blame follows that, whereas moral blame invites difficult questions of *which* morality is being imposed, reasonableness offers an altogether broader and more generally acceptable standard of conduct that can encompass and tolerate competing moral values within a modern and tolerant society. But, without reference to some moral code, how is reasonableness to be determined? The Supreme Court has given recent consideration to the concept of the reasonable man and, in a passage that bears repeating, describes how the concept is to be approached:

> 'The Clapham omnibus has many passengers. The most venerable is the reasonable man, who was born during the reign of Victoria but remains in vigorous health. Amongst the other passengers are the right-thinking

[15] Jonathan Herring, *Medical Law and Ethics* (7th ed. Oxford University Press 2018), 108; citing *Bolam v Friern Hospital Management Committee* [1957] 2 All ER 118; *Maynard v West Midlands Regional Health Authority* [1985] 1 All ER 635.
[16] Police and Criminal Evidence Act 1984, s. 1(3).
[17] *Ibid*., s. 17(2).
[18] *Ibid*., s. 24(1).
[19] *R v Howe* [1987] AC 417, 447 – 448.
[20] *Attorney-General for Northern Ireland's Reference (No. 1 of 1975)* [1977] AC 105, 137; Criminal Justice and Immigration Act 2008, s. 76.

> member of society, familiar from the law of defamation, the officious bystander, the reasonable parent, the reasonable landlord, and the fair-minded and informed observer, all of whom have had season tickets for many years… But its most famous passenger, and the others I have mentioned, are legal fictions. They belong to an intellectual tradition of defining a legal standard by reference to a hypothetical person, which stretches back to the creation by Roman jurists of the figure of the *bonus pater familias*…
>
> 'It follows from the nature of the reasonable man, as a means of describing a standard applied by the court, that it would be misconceived for a party to seek to lead evidence from actual passengers on the Clapham omnibus as to how they would have acted in a given situation or what they would have foreseen, in order to establish how the reasonable man would have acted or what he would have foreseen. Even if the party offered to prove that his witnesses were reasonable men, the evidence would be beside the point. The behaviour of the reasonable man is not established by the evidence of witnesses, but by the application of a legal standard by the court. The court may require to be informed by evidence of circumstances which bear on its application of the standard of the reasonable man in any particular case; but it is then for the court to determine the outcome, in those circumstances, of applying that impersonal standard.'[21]

This passage reflects both the origins of the principle traceable to Roman law alongside a number of further instances where reasonableness is the defining standard of conduct. Further, the passage describes how the concept is accessed and applied, as an objective legal fiction rather than the subject of witness evidence. As Lord Radcliffe provides more succinctly, 'the spokesman of the fair and reasonable man, who represents after all no more than an anthropomorphic conception of justice, is and must be the court itself.'[22] That being said, within the context of the criminal law it is not only the judge but, more

[21] *Healthcare at Home Ltd. v The Common Services Agency* [2014] UKSC 49, [1] – [3].
[22] *Davis Contractors Ltd. v Fareham Urban District Council* [1956] AC 696, 728.

so, the jury whom play a pivotal role in assessing the defendant's conduct. In this respect, reasonableness is somewhat democratised within the criminal court, as the concept will be considered and applied by twelve individual members of the jury, alongside the judge.

This contrast between passing moral judgment on the one hand and determining the boundaries of reasonable conduct on the other is roundly enunciated in *Re T (Minors) (Custody: Religious Upbringing)*,[23] which concerned a dispute for custody over children between parents with conflicting religious and moral views. Lord Scarman wrote:

> 'We live in a tolerant society. There is no reason at all why the mother should not espouse the beliefs and practice of Jehovah's Witnesses… It is as *reasonable* on the part of the mother that she should wish to teach her children the beliefs and practice of the Jehovah's Witnesses as it is *reasonable* on the part of the father that they should not be taught those practices and beliefs. It is not for this court, in society as at present constituted, to pass any judgment on the beliefs of the mother or on the beliefs of the father. It is sufficient for this court that it should recognize that each is entitled to his or her own beliefs and way of life, and that the two opposing ways of life considered in this case are both socially acceptable and certainly consistent with a decent and respectable life.'[24]

This judgment reflects two critical points submitted in the present section of the thesis. First, that the law generally, and the courts specifically, are not concerned with enforcing any particular moral code or passing moral judgment on the views, beliefs and lifestyles of others. Second, the judgment reflects the greater tolerance that is encapsulated by the concept of adjudging the reasonableness of conduct. That is to say, conduct that might regarded as being highly immoral by one group can nonetheless be accepted as being reasonable, having regard to the tolerance that must be given to opposing or contradictory views, beliefs and lifestyles. It is submitted that the standard of reasonableness, rather

[23] *Re T (Minors) (Custody: Religious Upbringing)* (1981) 2 FLR 239.
[24] *Ibid.*, 244 – 245 (emphasis added).

than moral blame, provides a more accurate and reliable foundation for the criminal law and legal responsibility.

The third argument in favour of reasonableness over moral blame follows that reasonableness provides a more coherent account of why the actions subject to criminal law and different criminal standards change over time. It is difficult to argue that homosexuality was fundamentally immoral conduct one day but not the next, or that actions such as rape within a marriage was morally acceptable one day but not the next. Accepting morality and moral blame as an underlying principle behind the criminal law thus requires accepting an extreme form of moral relativism that would be unacceptable to much of society and offers little guidance as to what should actually be criminalised. Conversely, it is far more readily arguable that the criminal law ultimately reflects the boundaries of what is considered by a society to be reasonable and unreasonable conduct. From this perspective, of course the law is bound to change over time as the attitudes of society towards different things equally evolve and change; such is the progress of civilisation. Nonetheless, this does not require accepting any degree of relativism with regards to what is *morally* good or bad, because this is not what the criminal law is fundamentally concerned with.

As a final point of interest, examples of movement away from notions of moral blame in favour of orientation towards a standard of socially reasonable behaviour may be found emerging in a number of jurisdictions around the world. For example, Papachristos, Meares and Fagan suggest that both Germany and Portugal have adopted an objective of 'positive general prevention' under which the 'primary purpose of criminal sanctions is the public reaffirmation of the validity of basic social norms that have been violated by the offender's flagrant norm violation.'[25] Similarly, Finland has moved from a focus of deterrence through harsher sentencing to one of general prevention whereby the 'norms of criminal law and the value they reflect are internalized.'[26] Whereas each jurisdiction

[25] Andrew V. Papachristos, Tracy L. Meares and Jeffrey Fagan, 'Why do criminals obey the law? The influence of legitimacy and social networks on active gun offenders' (2012) 102(2) *Journal of Criminal Law and Criminology* 397, 178; citing Thomas Weigend, 'Sentencing and punishment in Germany' in Tonry M. and Frase R. S. (eds.), *Sentencing and Sanctions in Western Countries* (Oxford University Press 2001), 209.
[26] Papachristos, Meares and Fagan (2012), 178; citing Tapio Lappi-Seppälä, 'The fall of the Finnish prison population' (2000) 1(1) *Journal of Scandanavian Studies in Criminology and Crime Prevention* 27, 28.

has naturally taken different routes, the key unifying feature is a focus of the criminal law on reflecting social norms of behaviour – what in this thesis is being referred to simply as "reasonableness" of behaviour. Crucially, this represents an abandonment of moral blame as the underlying justification for the criminal justice system.

*

Two key points may be made which bridge the gap between the principle of reasonableness underlying the criminal law and the various different formulations of *mens rea* considered in the next section, below. First, notwithstanding the relative familiarity that the courts have with applying standards of reasonableness across numerous areas of law and in countless different situations, "reasonableness" *per se* is nonetheless a somewhat vague and imprecise concept (albeit arguably no more so than the concept of moral blameworthiness). It is here that the different formulations of *mens rea* – intention, recklessness, knowledge, belief *etc.* – add particular value, for it is this which more specifically describes the circumstances in which a given action will be considered so unreasonable as to be criminal. For example, the requirement of an intention to kill or cause grievous bodily harm for the offence of murder denotes that it is criminally unreasonable to intentionally kill another; this principally separates the offences of murder from certain forms of manslaughter, highlighting the difference in "reasonableness", as it were, between killing intentionally (murder) and accidentally (manslaughter).

Similarly, the requirement of dishonesty for the offence of theft denotes that it is criminally unreasonable to take another's property in circumstances that are regarded as dishonest; again, this requirement could differentiate criminal theft from mistakenly picking up somebody else's property. In each respect, the different particular formulations of *mens rea* fulfil a dual function, on the one hand enabling Parliament to specify more precisely the type of conduct that it intends to criminalise and, on the other hand, providing more specific boundaries within which the courts may assess the particular responsibility of individual defendants on a case-by-case basis. It is important to note that these functions of the different formulations of *mens rea* do not necessarily

require that those formulations relate to *subjective* states of mind; the same ends may be achieved through objective formulations, such as those proposed in this thesis.

Second, Hart can provide further inspiration for the link between different forms of *mens rea* and the proposed underlying principle of reasonableness. Mirroring similar comments throughout this thesis and the preceding chapter in particular, Hart first writes that, from a deterministic viewpoint:

> '[I]t is always false, if not senseless, to say that a criminal could have helped doing what he did. So on this theory, when we inquire into the mental state of the accused, we do not do so to answer the question, Could he help it? Nor of course to answer the question, Could the threat of punishment have been effective in his case? – for we know that it was not.'[27]

Hart proceeds to draw an analogy between defences that excuse a defendant's criminal conduct – namely mistake, accident, coercion, duress and insanity – and comparable conditions that may invalidate civil transactions such as wills, gifts, contracts and marriages, including conditions of insanity, mistake, duress and coercion.[28] The point of the analogy with regards to both civil and criminal law is that these similar excusing and invalidating conditions reflect that fact that the defendant (or civil respondent) could not execute a real choice:

> '[T]he individual might have chosen one course of events and by the transaction procured another (case of mistake, ignorance, *etc.*), or he might have chosen to enter the transaction without coolly and calmly thinking out what he wanted (undue influence), or he might have been subjected to the threats of another who had imposed *his* choices (coercion).'[29]

---

[27] H. L. A. Hart, 'Legal responsibility and excuses' in Hart H. L. A. (ed.), *Punishment and Responsibility: Essays in the Philosophy of Law* (2nd ed. Oxford University Press 2008), 42.
[28] *Ibid*, 34 & 45 – 48.
[29] *Ibid*., 45.

Finally, with this in mind, Hart proposes that the function of exploring a defendant's mental state is therefore to maximise the criminal law's coercive deterrent effect on people's informed and considered choices. This is achieved in a threefold manner:

> 'First, we maximize the individual's power at any time to predict the likelihood that the sanctions of the criminal law will be applied to him.'[30]

This may be reasoned, for example, because the different formulations of *mens rea* spell out the specific circumstances in which a defendant's conduct would be regarded by the law as being criminal.

> 'Secondly, we introduce the individual's choice as one of the operative factors determining whether or not these sanctions shall be applied to him. He can weigh the cost to him of obeying the law – and of sacrificing some satisfaction in order to obey – against obtaining that satisfaction at the cost of paying "the penalty".'[31]

This reflects the undisputed assertion that people do make efficacious decisions, regardless of the metaphysical truth of determinism. People are constantly faced with scenarios where they must make a choice and, as has been explored in chapter seven of this thesis, people are capable of applying deliberation and reasoning to their decision-making processes (even if this does not provide online *conscious* control of that decision outcome). Decisions are, therefore, both meaningful and responsive to reason, and one of the principal aims of the criminal law is to therefore provide proscriptions against unreasonable behaviour and the penalties attached thereto, in order for these to become factors under consideration within people's decision-making. This is related to the last point:

> 'Thirdly, by adopting this system of attaching excusing conditions we provide that, if the sanctions of the criminal law are applied, the pains of

[30] *Ibid*., 47.
[31] *Ibid*.

> punishment will for each individual represent the price of some satisfaction obtained from breach of the law.'[32]

This, it is submitted, reflects one of the aforementioned distinctions between morality and law; whereas morality seeks to prescribe what people ought or ought not to do – (and thus, within the context of the present thesis, suffers from the challenge of reconciling determinism with the principle of *ought implies can*) – the law instead proscribes what people must not do *alongside* announcing the enforceable penalties for breach. Thus, defendants are forewarned of the consequences should they commit certain prohibited actions within specified formulations of *mens rea*. The forewarned consequences attached to legal proscriptions (but not moral duties) can therefore occupy a stronger causative role within an individual's decision-making which moral obligations (absent of the threat of coercive punishment) cannot similarly achieve, at least not to the same degree.

## *9.2. Hybrid Objective / Subjective Mens Rea*

The previous chapter of this thesis presented the various reasons why the evidence considered in Part One of the thesis is deemed to undermine subjective *mens rea*, providing an insecure foundation upon which to determine criminal liability. Instead, it is here proposed that *mens rea* is reformulated into a hybrid objective / subjective concept. On the one hand, the different formulations of *mens rea* such as intention and recklessness *etc.* will be given entirely objective definitions; on the other hand, however, it is proposed that the objective formulations of *mens rea* are applied to the hypothetical "reasonable defendant" who possesses the same subjective characteristics as the particular defendant in a given case. Inspiration for this hybrid objective / subjective approach is, once again, drawn from the work of H. L. A. Hart, and in particular his discussion of fault in offences of criminally negligent conduct. To begin, Hart distinguishes negligent liability from strict (or absolute) liability, writing that:

> '[A]bsolute liability results, not from the admission of the principle that one who has been grossly negligent is criminally responsible for the

[32] *Ibid*.

> consequent harm even if "he had no idea in his mind of harm to anyone," but from the refusal in the application of this principle *to consider the capacities* of an individual who has fallen below the standard of care.'[33]

In this passage, Hart is describing why it is justified to attribute criminal liability for negligent conduct, when the very definition of negligence is not at all dependent upon the defendant's subjective state of mind – *i.e.,* the defendant was neither minded to cause harm nor had foreseen that any such harm might occur. Hart's solution is that the standard of negligence – which is entirely objective – must be applied with the defendant's subjective capacities in mind. It is this similar approach that is proposed in the present thesis to be applied across *mens rea* entirely, with objective formulations of different *mens rea* applied taking each defendant's subjective capacities into consideration. Stewart breaks down Hart's approach into three distinct components: '(i) the reasonable person would have observed the appropriate standard of care; (ii) the accused departed markedly from that standard of care; and (iii) the accused could have met (had the capacity to meet) the standard of care.'[34] These may be broadly reflected onto the new hybrid objective / subjective approach to *men rea*.

Thus, (i) each different formulation of *mens rea*, in conjunction with *actus reus*, denotes the appropriate standard of conduct that everybody is reasonably expected to adhere to, the breach of which is regarded as being criminally unreasonable conduct; (ii) the defendant departs from that standard of reasonable conduct whereby, following the objective formulations of *mens rea*, such departure 'may be externally observed';[35] and (iii) the defendant had the capacity to act within that standard of conduct, which is to say that it is reasonable to expect any other person in the same subjective circumstances as the particular defendant to have complied with that standard of conduct.

---

[33] H. L. A. Hart, 'Negligence, *mens rea*, and criminal responsibility' in Hart H. L. A. (ed.), *Punishment and Responsibility: Essays in the Philosophy of Law* (2nd ed. Oxford University Press 2008), 154 – 155.
[34] Hamish Stewart, 'Legality and morality in H. L. A. Hart's theory of criminal law' (1999) 52(1) *SMU Law Review* 201, 208.
[35] *Ibid*.

The first limb of this conceptualisation of *mens rea* may be read as reflecting the idea that the criminal law is concerned with reasonable and unreasonable conduct as opposed to moral blame, discussed above. The second limb reflects the fact that the defendant has acted (or has been externally observed as acting) unreasonably, as objectively defined by *mens rea*. The third limb may be read as reflecting three capacities deemed necessary for responsibility: a revised presumption of volition that focuses on a capacity for being responsive to reason and a capacity for "ordinary" self-control, (rather than any notion of *conscious* control); and the capacity to appreciate the nature and consequences of one's actions. Where one of more of these three capacities are absent, an individual may be regarded as lacking the overall capacity to meet the requisite criminal standard such that it is not reasonable to expect anybody in the same circumstances to have acted differently.

Does this hybrid objective / subjective approach to *mens rea* resolve the issues raised in chapter eight, above? The discussion of the *what*, *when* and *whether* components presented a significant challenge to the assumption (and common intuition) that people have conscious control over their decisions and actions; that is to say, whilst people obviously do have (and generally exercise) a capacity for self-control over their behaviour and actions, it is refuted that this is a capacity of (or is caused by) consciousness, nor that this capacity enables control over the actual outcome of automatic decision-making processes in the brain. This undermines a key presumption behind subjective *mens rea*, namely that people are responsible for their actions because they have the conscious ability to control their decisions and act otherwise.

Under the hybrid objective / subjective approach, however, whilst *mens rea* might still colloquially be understood as describing states of mind, the objective descriptions of those mental states are more akin to *actus reus* in as much as the offending criminal behaviour is observed externally and assessed against objective criteria. Meanwhile, the application of these objective *mens rea* to the defendant's particular subjective circumstances – in particular in so far as those circumstances relate to a person's capacity appreciate that nature and consequences of their conduct, to act according to reason and with self-control (*i.e.,* the capacity to act within a criminally reasonable standard of conduct) – is what distinguishes objective / subjective *mens rea* from absolute liability.

The critical relation is that of objectively defined mental states to any such circumstances of the defendant that impact upon their relevant capacities. As Stewart writes in relation to Hart and criminal negligence, 'fault is thus located not simply in the departure from the relevant standard of care' – *i.e.,* objectively defined *men rea* – 'but in departing from that standard *even though the accused was capable of meeting it*.'[36] It must be stressed, the fact that objective *mens rea* is proposed to be applied in light of a defendant's subjective circumstances is not an indirect or circular return to subjective *men rea*. Whilst evidence of a defendant's subjective state of mind would undoubtedly remain probative to establishing whether or not they had breached an objectively defined standard, proof of any such subjective state of mind would not be a necessary component of *mens rea*. Equally, proof of "objective" circumstances which impact upon a person's capacity for reasons responsiveness, ordinary self-control, and appreciating the nature of their actions, would be inherently relevant in applying objective *mens rea* to the particular circumstances of each defendant's case – such circumstances would include, for example, age and mental maturity, illness or disability affecting rational capabilities, and addiction.

Discussing a general move by Parliament to increasingly introduce objective forms of *mens rea*, Andrew Ashworth and Jeremy Horder[37] advocate for an approach that is not too dissimilar to the hybrid approach proposed in this thesis. In particular, they explain that objective tests ought to be applied subject to exceptions based upon the defendant's capacities, arguing that this 'respects the principle of moral autonomy, by ensuring that no person is convicted who lacked the capacity to conform his or her behaviour to the standard required.'[38] The hybrid approach to *mens rea* achieves this through the two-limb test. For each type of *mens rea*, the objective first limb of the test asks whether the defendant's conduct meets an objective description of the particular *mens rea* concerned. The subjective second limb of the test then asks whether or not it is reasonable to expect anybody sharing the defendant's subjective characteristics and circumstances (*i.e.,* the "reasonable man" imbued with relevant features of the particular defendant) to appreciate how their actions relate to the *mens rea* of the offence as objectively defined. The

[36] *Ibid*., 208 (emphasis added).
[37] Andrew Ashworth and Jeremy Horder, *Principles of Criminal Law* (7th ed. Oxford University Press 2013), 185; Jeremy Horder, *Ashworth's Principles of Criminal Law* (9th ed. Oxford University Press 2019), 208.
[38] *Ibid*.

subjective circumstances and the characteristics that the defendant may adduce under this reasonableness test are those that may be relevant to the capacity to so appreciate the nature and consequences of their actions, or to their presumed volitional capacities of being responsive to reason and exercising ordinary self-control.

In addition, the discussion of the *why* component of decision-making revealed that, whilst on the one hand people are generally adept at reasoning socially and through argumentation and, indeed, are often responsive to reasons, people also generally have a poor subjective access to genuine reasons for actions. This is explained by suggesting that the capacity for reasoning and rational thought operates *post hoc*, providing justification for decisions that have been reached through more intuitive processes, rather than providing foundational reasons upon which decisions are based. However, it is the poor subjective access to genuine reasons that is of particular interest; people appear generally inept at distinguishing between genuine reasons and confabulations. Notwithstanding the challenges faced by a court in identifying dishonest testimony, it must also identify testimony that is honestly confabulated, that the defendant themselves cannot distinguish from genuine reason. This raises further significant challenges in relying upon subjective mental states as a basis for legal responsibility; the court faces a fundamental challenge in ever being able to verify something (a state of mind) that is entirely subjective.

Under the proposed hybrid approach, however, proof of subjective states of mind may remain evidentiary but are not probative of legal responsibility. As mental capacities are considerably more stable and evidenced through a whole range of observable behaviour in general, proof or disproof of their existence can be obtained with considerably greater reliability *and* from the objective perspective of the courtroom. Meanwhile, the court will continue to have regard to relevant subjective characteristics and circumstances of each individual defendant which demonstrably impact upon these capacities, thereby ensuring fairness by holding defendants to a standard of conduct that remains achievable.

### *9.2.1. The Example of Intention*

Undoubtedly, the best demonstration of the proposed hybrid objective / subjective approach to *mens rea* is through example which, for present purposes, may be achieved by exploring the *mens rea* of intention. This section does not enter into a full exploration of the hybrid approach to *mens rea* in relation to intention specifically, which is considered in greater detail in section 10.1 of this thesis, below. Rather, the present discussion serves simply to illustrate the broader hybrid approach to *mens rea* that is being advocated in this thesis.

First, intention may be defined objectively as existing when the prohibited outcome (*e.g.,* death or injury to another) was a 'virtual certainty (barring some unforeseen intervention) as a result of the defendant's actions.'[39] This formulation is borrowed from the concept of oblique intention in *R v Woollin*, which is explored in greater detail in section 10.1.1, below. Thus, the first question is an objective inquiry regarding the virtual certainty with which the prohibited outcome was likely to follow from the defendant's conduct. The second question introduces subjective elements by reference to a particular defendant's circumstances as well as incorporating the concept of reasonableness which underlies *mens rea*. The second question asks, *is it reasonable to expect anybody in the defendant's circumstances to appreciate that virtual certainty?* Thus, on the one hand, the second question is making direct reference to the concept of reasonableness, inviting the court to assess whether the defendant's conduct meets the standard of reasonable conduct expected of anybody in society. On the other hand, however, the second question also invites consideration of the defendant's subjective circumstances, so far as those circumstances relate to their capacity to meet that standard of reasonable conduct.

The first, objective definition of each type of *mens rea* is relatively straight forward; however, the second subjective element requires further elucidation. Expressed in full, the question asks *whether it is reasonable to expect anybody in the defendant's (subjective) circumstances to appreciate that the prohibited outcome (e.g., killing or injuring another etc.) was virtually certain to result from the defendant's actions*. The subjective circumstances that are relevant to this question will be those that bear any relation to the

[39] *R v Woollin* [1999] 1 AC 82, 96; citing *R v Nedrick* [1986] 1 WLR 1025, 1028.

capacity for anybody to appreciate the virtual certainty of the prohibited consequences following particular actions, and any relation to the volitional capacities for rational thought and exercising self-control in general, the latter of which forms the revised presumption of voluntariness discussed in the following section of this chapter, below. Of course, the defendant must still prove those subjective circumstances that are claimed before the court, which is one reason why inquiries into the subjective state of mind of a defendant will remain valuable in providing evidence of those circumstances.

However, what renders such circumstances as being *relevant* is the impact that those circumstances would have on *anybody's* capacities as described above. The capacity to appreciate the virtual certainty of consequences in particular might be negated, for example, through evidence that has traditionally been used to negate subjective *mens rea*, such as claiming mistake as to relevant facts or claiming to have acted accidentally. The capacities for rational thought and self-control more generally relate to the revised presumption of voluntariness underlying *mens rea* and might be negated, for example, through evidence that has traditionally been the subject of defences such as duress, necessity and self-defence. Crucially, by reference to circumstances that would impact upon *anybody's* capacities, the defendant's conduct in any particular case is still being compared to an objective standard of reasonable conduct, albeit that objective standard has been formulated with the defendant's relevant subjective circumstances in contemplation.

Consider a hypothetical example of potential theft in which a female defendant picks up another woman's handbag on the metro and walks away. Theft is defined as the dishonest appropriation of property belonging to another with the intention of permanently depriving the other of it;[40] for present purposes, the *mens rea* of dishonesty will be disregarded and the focus shall be on intention alone. Applying the revised hybrid formulation, above, to the present case, the defendant will be *prima facie* guilty of theft if: (a) it is a virtual certainty (barring some unforeseen intervention) that, by picking up another woman's bag and walking away, she would be permanently depriving that woman of her property; and (b) it is reasonable to expect that anybody in the defendant's

[40] Theft Act 1968, s. 1(1).

circumstances would appreciate that virtual certainty. Part (a) may be established relatively easily from the scant facts of the present case; the defendant has picked up somebody else's bag on the metro and walked away. Assuming an ordinary course of affairs, the metro has proceeded to drive off whilst the defendant has walked away with another woman's bag; barring any further specific action to reunite the two, the victim has been permanently deprived of her property.

Part (b) invites further exploration, the vast majority of which is ordinarily the subject of a defendant's submissions to negate subjective *mens rea* or to establish a legal defence. For example, suppose the woman's handbag is identical to the defendant's, and the defendant pleads that they simply picked up the wrong bag by mistake. This is a mistake as to fact and, provided that the defendant's testimony is credible and believed – perhaps with the support of evidence adducing the similarity of the two handbags – this would be a relevant circumstance for the defence to raise. If established successfully then the answer to part (b) is an emphatic *no*; it is not reasonable to expect that somebody who picks up what appears to be their bag, believing it to be their own, *would appreciate that they are virtually certain* to permanently deprive somebody else of their property.

Suppose the defendant knew that the bag was not their own, but that they were obliged to steal it on a secret mission for the CIA. Alongside psychiatric evaluation, this could provide evidence of schizophrenia in order to establish the defence of temporary insanity (assuming the defendant is not actually a spy!). If established successfully then the answer to part (b) is also, arguably, a *no*; it is not reasonable to expect somebody suffering from schizophrenia to necessarily appreciate the virtually certain consequences of their actions. In addition, this defence would also speak to the underlying presumption of voluntariness, as schizophrenia and other mental illnesses can clearly have an impact on people's capacity for rational thought and self-control.

Consider an alternative hypothetical case of potential assault and battery in which a male defendant's hand reaches out and forcibly hits another man on the moving metro. For present purposes, let assault and battery be defined as the intentional or reckless

application of unlawful force upon another;[41] and, again, focus shall be solely on the *mens rea* of intention. Following the revised formulation, the defendant will be *prima facie* guilty of assault and battery if: (a) it is a virtual certainty that, by reaching out their hand towards another on the metro, he would forcibly strike that other person; and (b) it is reasonable to expect that anybody in the defendant's circumstances would appreciate that virtual certainty. Again, part (a) presents an objective description of the *mens rea* of intention and, again assuming an ordinary course of affairs, it is virtually certain that if a person reaches out their hand swiftly towards another person in proximity, then they may strike that other person with force. Part (b) invites further investigation.

Suppose that the metro is crowded and in motion, and that the defendant lost their balance when they suddenly reached their hand out towards another person; they were not reaching for that person but for a nearby handrail, and simply struck the victim by accident. In legal terms, this would be an argument simply to negate the existence of subjective intention; the assault and battery was committed by accident and without the requisite intention. If indeed the defendant credibly establishes that they lost their balance and were reaching for a nearby handrail, part (b) will likely be answered in the negative; it is not reasonable to expect that somebody reaching out for a handrail would appreciate that they are virtually certain to forcibly strike another person. It might be reasonable to expect such a person to appreciate the possibility, or even the likelihood, of striking another whilst reaching for stability on a busy moving metro; but appreciating a *virtual certainty* is a different matter.

Suppose, instead, that the defendant is diabetic and suffers from hypoglycaemia, resulting in his involuntarily reaching out his arm and striking the victim; this *prima facie* describes a defence of automatism under which the defendant was not in control of their actions as a result of some medical condition. Considering again part (b) of the hybrid formulation of intention, it is not reasonable to expect that somebody suffering from automatism induced by hypoglycaemia would appreciate that they are virtually certain to forcibly strike another, precisely because anybody in such a situation would be unaware of the action of reaching out their hand. Furthermore, the defence of automatism would also

[41] *Director of Public Prosecutions v Little* [1992] QB 645; Criminal Justice Act 1988, s. 39.

relate to the presumption of voluntariness, as automatism would negate anybody's capacity to respond to reason and exert ordinary self-control over their actions.

These examples are by no means exhaustive; chapter ten of this thesis, below, provides a more detailed elaboration of each type of *mens rea*, whilst chapter eleven proceeds to consider the interaction of legal defences. That notwithstanding, the examples provided here demonstrate how an objective description of *mens rea* such as intention can be applied, whilst taking into consideration relevant subjective circumstances of a particular defendant. The circumstances that are relevant to adduce will be those that relate to either of the three crucial capacities for responsibility; the capacity for anybody in the defendant's circumstances to reasonably appreciate the specific *mens rea* under consideration, and the general capacities to be responsive to reasons and to exert ordinary self-control. The hybrid objective / subjective formulation of *mens rea* encapsulates the underlying principle that *mens rea* in itself reflects criminally unreasonable conduct, by inviting an inquiry as to whether the defendant's objectively defined conduct breached the standard of reasonable behaviour expected of anybody (*i.e.,* the hypothetical "reasonable man") possessed of the defendant's relevant subjective circumstances.

## *9.3. Rational Thought and Ordinary Control*

Discussed in section 8.1, above, the legal presumption of volition includes two assumptions; first, that people have the capacity for rational thought and reasons responsiveness and, second, that people have a capacity for conscious online control over their decisions and actions. On the one hand, the evidence considered in chapter seven of this thesis provides support for the first of these assumptions. People evidently do have the capacity to be adept at rational thought and are responsive to reasons, in particular where this takes places within social contexts of debate and argumentation. The evidence suggests that views, opinions and judgments are formed more intuitively, in the sense that they are determined through unconscious mental processes and then presented to conscious awareness as conclusions, rather than being established and reasoned from first principles. That notwithstanding, the evidence equally suggests that reasoning and

argumentation can impact upon and change these judgments; reason can therefore exert a causal effect on people's decisions such that people are indeed reasons responsive.

The second assumption requires some reconsideration, however. In particular, the evidence across Part One of this thesis strongly suggests that consciousness *per se* does not provide any source of control over decisions and actions. Rather, self-control exists independently of consciousness; it is something that is developed from the early years of childhood, and exists as one example of what are broadly referred to as executive functions. It is even arguable that self-control is itself a prerequisite for conscious deliberation, rather than a product of conscious thought. In order to consciously deliberate and evaluate a number of competing decision alternatives, an individual must possess the necessary self-control to avoid simply selecting the first option that comes to mind or reaches a particular degree of valence. Conscious deliberation thus *first* requires a degree of self-control, and not *vice versa*. The absence of this can be seen in children who are yet to develop sufficient self-control. Instructed, for example, to select only one toy in a toy store, most young children will pick up the first toy that they see, and then likely pick up more and more as they proceed around the store. They do not yet possess the requisite self-control to inhibit their impulse to select the first attractive thing and spend more time deliberating on what might be their preferred choice.

The presumption of voluntariness can be amended relatively easily in light of these considerations. Simply, the presumption of voluntariness assumes that people have the capacity for rational thought and reasons responsiveness, and the capacity for *ordinary* self-control necessary for such rational thought and for conforming bodily actions to intentions or goals. The only statement that needs to be abrogated is that which relates the notion of self-control specifically to consciousness. Discussing the separation of consciousness from volition from a neuroscientific perspective, Bonn writes:

> 'Strictly conscious control over behaviour seems to be ruled out by our improved understanding of the mind. Does this mean, however, that a person is not in control of their behaviour? Again, keeping in mind a broad definition of the "person" as including both conscious and unconscious

> elements, recent discoveries can shed light on this issue. First, there is a separate motor control network dedicated to internally generated, voluntary, goal-oriented behaviours as contrasted with externally-triggered and more habitual behaviour. Second, there appear to be connections between the default network, where novel ideas and counterfactual scenarios are produced, and this goal-oriented control network that allows for the internal generation of action.'[42]

The crucial point being captured here is that the brain possesses mechanisms that can produce novel, internally generated goals and intentions and, furthermore, possesses mechanisms for controlling actions in order to pursue those goals and intentions – *i.e.,* people clearly possess self-control necessary for goal-oriented behaviour. As Bonn explains, consciousness itself appears to have a minimal input into these underlying processes of self-control; however, this does not mean that people do not act volitionally, only that the understanding of volition requires revision. Indeed, this idea that self-control is something altogether more unconscious and automatic is in concurrence with the broader point asserted by many of the commentators discussed throughout this thesis, that '*most* of our behavioural responses are, in fact, automatic'[43] or that, taking the assertion to its most extreme, 'conscious experiences of will do not cause human actions at all.'[44]

The alteration of volition from reasons responsiveness and conscious control to reasons responsiveness and *ordinary* self-control – or, more elaborately, the requisite degree of ordinary self-control necessary to engage in rational thought and reasons responsiveness – may appear superficial or inconsequential. Conversely, it is submitted that this revision

[42] Gregory B. Bonn, 'Re-conceptualizing free will for the 21st century: Acting independently with a limited role for consciousness' (2013) 4 *Frontiers in Psychology* 1, 5.
[43] Natalie S. Gordon and Mark R. Fondacaro, 'Rethinking the voluntary act requirement: Implications from neuroscience and behavioral science research' (2018) 36(4) *Behavioral Sciences & the Law* 426, 433; citing Benjamin Libet, 'Unconscious cerebral initiative and the role of conscious will in voluntary action' (1985) 8(4) *Behavioral and Brain Sciences* 529; John A. Bargh and Tanya L. Chartrand, 'The unbearable automaticity of being' (1999) 54(7) *American Psychologist* 462; Daniel M. Wegner and Thalia Wheatley, 'Apparent mental causation: Sources of the experience of will' (1999) 54(7) *American Psychologist* 480; Roy F. Baumeister, E. J. Masciampo and Kathleen D. Vohs, 'Do conscious thoughts cause behavoir?' (2011) 62(1) *Annual Review of Psychology* 331; Michael S. Gazzaniga, *Who's in Charge? Free Will and the Science of the Brain* (Robinson 2012).
[44] *Ibid*; citing Daniel M. Wegner, *The Illusion of Conscious Will* (Massachusetts Institute of Technology Press 2018), 318.

is significant for at least two reasons, each articulated by Gordon and Fondacaro.[45] First, this helps reconceptualise responsibility 'as a socially constructed rule or obligation that emerges out of the interactions among individuals in a social context rather than as an internal individual attribute.'[46] Of course, "ordinary self-control" is every bit an internal attribute as conscious control; however, conscious control implies that self-control is something more than largely automatic, and if often conflated with philosophical attributions of free will.[47] Second (and perhaps more convincingly), the assumption that people have conscious online control over decisions and actions is a principal justification for retributive theories of justice that are otherwise rejected by arguments throughout this thesis; this latter point is explored more fully in section 12.1 of this thesis, below.

The revised conception of volition as reasons responsiveness with ordinary self-control is reflected in the second limb of the hybrid formulation of *mens rea*; thus, whereas the first limb provides an objective definition of any given type of *mens rea*, the second limb asks *whether it is reasonable to expect anybody in the defendant's circumstances to appreciate that objective mens rea, i.e., the nature and consequences of their actions*. As was discussed in the previous section, above, the various types of circumstances subjective to any given defendant would be those circumstances that impact upon anybody's capacity to appreciate the objective *mens rea* or be reasons responsive with requisite self-control (*i.e.,* volitional action). Those such circumstances that are relevant to volition in particular generally form the basis of most legal defences.

For example, self-defence impacts upon volition because it is recognised that most people would not be able to refrain from defending themselves against the threat of immediate violence. Similarly, the defences of duress and necessity recognise that people and circumstances respectively can exert such a coercive influence over an individual's responsiveness to other reasons or self-control that they no longer act volitionally. Equally, automatism and temporary insanity each clearly relate to circumstances (typically recognised medical or psychiatric disorders) that impact upon volition as herein

[45] Gordon and Fondacaro (2018).
[46] *Ibid*, 433.
[47] *Ibid*., 4 – 5.

defined. For example, somebody who commits the actions of a criminal offence whilst in a state of epileptic- or hypoglycaemic-induced automatism, *virtually by definition*, lacks the ordinary capacities to control their physical motions so that they correspond with intended bodily actions. Thus, the revised presumption of volition is encapsulated within hybrid reformulation of *mens rea*. The relation of these and other defences to the three crucial capacities for responsibility is the subject of chapter eleven of this thesis, below.

## *9.4. Linking Capacities to Responsibility*

The reconstruction of *mens rea* presented here proposes, in essence, that legal responsibility arises when a person commits a criminal act in circumstances where they were in possession of three mental capacities – the presumed volitional capacities of being capable of rational thought that responds to reason and of being able to exercise ordinary self-control over bodily actions; and the third capacity to appreciate the nature and consequences of one's actions, which must be proven by the prosecution under the hybrid objective / subjective approach. Why, though, should the presence of these three capacities be sufficient for legal responsibility? In particular, recalling the clinical case of a patient whose brain tumour demonstrably caused his deviant sexual behaviour presented at the introduction to this thesis, we are reluctant to hold somebody responsible for behaviour resulting clearly and inescapably from factors outside of their control. But, is it not the case that whether or not an individual possesses the three aforementioned capacities is equally a matter outside of their control? What is more, recalling the broader deterministic perspective in the introduction to this thesis, if all thoughts, desires and intentions are, again, ultimately caused by factors outside of the individual's control, so too must be the presence (or absence) and exercise of the aforementioned capacities.

It is submitted that the justification for ascribing responsibility in the presence of the three mental capacities is intrinsically linked with the fundamental purpose of any system of legal and / or moral rules and, perhaps more importantly, the legitimate responses to a breach of those rules that are necessary in order to uphold their fundamental purposes. Specifically, it is first proposed that any system of legal / moral rules functions to identify both desirous and devious behaviour, with a view to encouraging, persuading, coercing

and even compelling people to conform their own behaviour accordingly. The specific content of those rules is immaterial to the question of responsibility for their breach and will be particular to each individual community, society, or nation. Equally, any given system of legal / moral rules will invariably exist on a continuum, with the least serious consisting of rules of manners and etiquette and the most serious being the equivalent of criminal prohibitions, whilst the severity of responses to the breach of those rules will generally correspond with its position on this continuum.

Second, it is submitted that the three capacities of reasons responsiveness, self-control and appreciation of the nature and consequences of one's actions are entirely necessary and sufficient for any person to conform their behaviour to a given legal / moral rule. That is to say, anybody having these three capacities possesses all that they theoretically require to incorporate into their decision-making the fact that certain conduct is prohibited and that such prohibition is a good reason not to engage in that conduct; to exercise the necessary self-control to avoid engaging in that conduct; and to appreciate the nature of their own actions so that they can select those which do not breach the prohibition. Equally, the presence of these three capacities may be necessary in order for any individual to be appropriately responsive to criminal punishment and rehabilitation. With regards to this second proposition, the fact that whether or not an individual possesses the three capacities is itself a matter entirely determined and outside of their subjective control is immaterial to the fact that these capacities are alone theoretically necessary and sufficient to enable anybody to conform to a given legal / moral proscription.

Given the previous proposition and the underlying perspective of determinism, third, if a person in possession of everything necessary and sufficient to conform their behaviour to a set of legal / moral rules (*i.e.,* the three capacities) nevertheless decides to act in breach of those rules, there must necessarily exist some factor(s) which not only caused that person's prohibited conduct, but were sufficient to do so *notwithstanding the fact* that that conduct was prohibited. Indeed, this may simply be read as an expansion upon the basic underlying argument of determinism; all behaviour is caused, and the fact that certain conduct is proscribed by some legal / moral rule is itself a factor which should contribute to causing a person not to perform that prohibited behaviour – it is a prototypically *good*

*reason* for not committing the prohibited action. Consequently, a person possessing the requisite three capacities who engages in behaviour that prohibited conduct must have been caused to so behave by some other factor(s) which, at the very least, outweighed the fact of the behaviour's prohibition in the individual's decision-making process.

Fourth and finally, the overbearing factor(s) resulting in the proscribed behaviour must be addressed (so far is it is possible to do so), whereas the failure to do so would undermine the very purpose for which the prohibitive legal / moral rule exists under the first proposition, *i.e.,* to identify desirous and devious behaviour and *encourage, persuade, coerce or compel compliance therewith*. Equally, the failure to address the causes of that criminal behaviour or, if this cannot be achieved directly, the failure to take necessary steps to prevent the individual from being able to repeat that behaviour, renders the criminal conduct more liable to be repeated again in the future. The phrase "be addressed" is interpreted broadly here and refers to the various ways in which a system of legal / moral rules can respond to their breach. In terms of the criminal law, this broadly refers to the different approaches to punishment; thus, the causes and effects of criminal behaviour may "be addressed" by incapacitating the individual, through measures designed to specifically deter them from similar conduct in the future, or through rehabilitation intended to address the underlying causes of their behaviour, *etc.*

The aforementioned propositions enjoy a loose Aristotelean pedigree. In brief, Aristotle regards virtuous behaviour to be desirable whilst vicious behaviour is not, to which end legal / moral rules identify and prescribe that which is virtuous and proscribe that which is vicious. However, a necessary prerequisite of responsibility and subsequent punishment is that these concepts attach to *voluntary actions*, because it is only those actions which are voluntarily chosen which fall within the sphere of an individual's control whilst, crucially, voluntariness is required for an individual to be responsive to the persuasion of legal / moral rules and accompanying praise or punishment:[48]

[48] Christof Rapp, 'Free will, choice, and responsibility (Book III.1-5 [1-7])' in Höffe O. (ed.), *Aristotle's "Nicomachean Ethics"* (Brill 2010).

> 'For they punish and seek revenge on those who do corrupt things (insofar as the latter do not act as a result of force or on account of an ignorance of which they are not themselves the cause), and they honor those who do noble things, *on the grounds that they will thereby exhort the latter and punish the former*. And yet nobody exhorts us to do those things that are neither up to us nor voluntary, *on the grounds that it is pointless to persuade someone not to feel heat or suffer pain or be hungry or any other such thing, since we will suffer them nonetheless*.'[49]

This passage is open to some interpretation; for example, Speight submits that those concerned with extrapolating a distinctively moralistic conception of responsibilty might focus on the retrospective aspect of punishment attaching to things having already been done. However, as he proceeds to interpret, the passage also clearly contains a prospective and consequentialist notion of punishment that corrects and reforms future behaviour.[50] Churchland explains more equivocally, 'the main idea was that if punishment in a certain type of case would neither deter nor improve the future behaviour of a person, including the defendant – if the punishment in these circumstances fails to provide a reason to avoid the action in the future – then full responsibility does not apply.'[51]

As mentioned above, the proposition that the factor(s) causing an individual's criminal behaviour *must* be addressed otherwise the purpose of the criminal prohibition will be undermined links the issue of responsibility to the purpose of the criminal justice system itself and the legitimacy of its responses to prohibited conduct. To express these four propositions in direct relation to criminal responsibility, therefore: (1) criminal laws exist to identify conduct that has been prohibited by a society and compel people from engaging in that conduct; (2) a person possessing the three capacities has all that is necessary and sufficient in theory to conform their behaviour with the law; (3) a person who commits a criminal act whilst in possession of the three capacities must have been

[49] Aristotle, *Nicomachean Ethics* (Bartlett R. C. and Collins S. D. (trns.), University of Chicago Press 2011), 51 – 52.
[50] Allen Speight, '"Listening to reason": The role of persuasion in Aristotle's account of praise, blame, and the voluntary' (2005) 38(3) *Philosophy & Rhetoric* 213, 216.
[51] Patricia S. Churchland, 'Moral decision-making and the brain' in Illes J. (ed.), *Neuroethics: Defining the Issues in Theory, Practice, and Policy* (Oxford University Press 2006), 9.

caused to so act by factors which overwhelmed the fact of that action's illegality in their decision-making process; (4) the relevant criminal prohibition and its purpose of compelling behaviour would be undermined if those overwhelming factors could not be addressed through the imposition of responsibility and subsequent punishment. By imposing responsibility, the effects of such overwhelming factors can be abrogated, for example, by incapacitating the affected convict so that the continuation of those causative factors cannot elicit a repetition of criminal conduct; or similarly, those factors might be mitigated through rehabilitation so that they no longer cause criminal behaviour.

If these submissions are accepted, then it follows that an individual may be held responsible for their actions, not simply regardless of but because of the causally deterministic world within which they act and, more specifically, the causal processes which govern decision-making in the brain. Responsibility does not rest upon a single determinant such as a subjective mental state. Rather, responsibility arises because of the combination of a breach of some legal / moral rule which purpose is to guide or compel behaviour, and that breach being committed by a person possessing the necessary capacities to be so guided by that rule. Because it necessarily follows that the resultant criminal conduct was caused by some overwhelming factor(s), it is not unreasonable to suspect that similar criminal conduct will be repeated unless those factors are addressed. The criminal justice system can address those factors and their resultant criminal conduct, for example, by incapacitating or rehabilitating the particular individual, as well as deterring others from being swayed into similar criminal conduct.

Moreover, the failure of the criminal justice system to appropriately respond to the breach of some criminal prohibition will undermine its very purpose and effectiveness as a means of guiding and compelling conduct in the first place. Therefore, it is because of the combination of these points – the breach of a behaviour-guiding rule, the presence of the requisite capacities to conform with the rule, the necessary existence of overwhelming factors resulting in criminal conduct, and the necessity for the criminal justice system to respond accordingly or otherwise undermine the concept of criminal prohibition in the first place – that responsibility for decisions and actions can fairly and rationally be attributed, notwithstanding the truth of causal determinism and absence of free will.

# 10. Elaborating Hybrid Objective / Subjective *Mens Rea*

> 'To the extent that neuroscience becomes better and better at predicting what we will do without reference to our personal volition, it will be less and less appropriate to treat people as freely acting agents. Predestination will become part of our real world.'
>
> - Sean Carroll, 2016.[1]

Following from the explanation of the proposed hybrid objective / subjective approach to *mens rea* in the previous chapter of this thesis, the present chapter proceeds to elaborate that hybrid approach across each of the different forms of *mens rea*. Each discussion aims to show through jurisprudence, not only how the reformulated approach to *mens rea* can be adopted into the existing law but, furthermore, how the reformulated approach may settle a number of extant tensions in jurisprudence between entirely subjective and objective approaches to *mens rea*. Most of the forms of *mens rea* explored in this chapter have been the subject of extensive consideration and revision in jurisprudence throughout the decades, as difficult cases gave rise to the need to reconsider previous definitions of terms like intention and recklessness. These difficult cases will therefore provide valuable tests in order to investigate how the hybrid approach advocated in this thesis might have been applied in practice to real legal cases.

To offer a very general degree of categorisation, the various types of *mens rea* (as reformulated under the hybrid objective / subjective approach) might be considered thus:

- Crimes of *intention* and *recklessness* are predominantly concerned with the capacity to appreciate a likelihood or foreseeability of some (harmful)

[1] Sean Carroll, *The Big Picture: On the Origins of Life, Meaning and the Universe Itself* (Dutton 2016), 384.

consequence, referring respectively to actions that are *virtually certain* to result in, or carry an *unreasonable risk* of resulting in those consequences.

- Crimes of *knowledge, belief, suspicion* and *dishonesty* refer to the capacity to appreciate states of existence, *i.e.,* respectively concerning the appreciation of *certainty* as to particular facts, the *conviction* as to particular facts, the *conjecture* as to particular facts, and the *dishonest nature* of certain conduct.
- Crimes of *negligence* are predominantly concerned with the breach of certain legal duties below the standard of the reasonable man, and relates to the capacity to appreciate the *unreasonableness* of conduct which falls below this standard.

## *10.1. Intention*

Under the revised hybrid formulation, intention is defined objectively as occurring when *the prohibited criminal outcome was virtually certain (barring some unforeseen intervention) to result from the defendant's act*, and is assessed with the defendant's relevant subjective circumstances in mind by asking, *is it reasonable to expect anybody in the defendant's circumstances to appreciate that virtual certainty*?

The concept of "virtual certainty" is taken from the definition of oblique intention, explored below in section 10.1.1. As is clear from the terminology, *virtual* certainty is something marginally short of absolute or scientific certainty. In this respect, demanding *absolute* certainty of outcomes as a prerequisite for intentionality would arguably set the threshold too high. People often intend some particular (criminal) outcome through actions which are less than absolutely certain, but nonetheless highly likely, to successfully achieve that outcome. Requiring absolute certainty would therefore restrict the definition of intentionality only to those actions which have been so precisely and acutely calibrated so as to guarantee a particular outcome, thereby excluding a larger majority of less certain and precise, but no less intentional criminal acts. Conversely, the threshold for intentionality cannot be established too low, such as requiring only a likelihood that some particular outcome may result from given actions. Such a lower threshold would blur any boundary between intention and recklessness, where intention is typically reserved for more serious offences such as murder, whilst less serious offences

such as manslaughter may be established through recklessness. Furthermore, the difference between offences which must be committed intentionally and those which may be committed recklessness forms the distinction between offences of specific and basic intent respectively in English law. This distinction becomes crucial when considering intoxication defences, discussed further in section 11.3.3 of this thesis, below.

The objective definition of intention includes the further clarification that a prohibited criminal outcome must be virtually certain to follow from a defendant's actions *barring some unforeseen intervention*. This addition serves to indicate that a prohibited outcome may still be regarded as objectively intentional, notwithstanding that some further unforeseen intervention prevented a given act from producing a prohibited outcome which was otherwise virtually certain to have followed in the natural course of events. In this sense, the term "unforeseen" should not be interpreted too strictly. For example, where a defendant intends to kill another and successfully shoots their target, but that target subsequently receives medical treatment, this does not extinguish the defendant's criminal intention. A victim receiving medical treatment in such circumstances is not entirely unforeseen in the true sense of the word. Rather, the clarification means that the victim's medical treatment was not a natural, inevitable, or even entirely obvious intervention to have occurred following the defendant's actions – medical assistance might never have been called, might have arrived too late, or might have otherwise been unsuccessful in saving the victim. The point is that, *without* the occurrence of some further intervention (which is itself neither guaranteed nor entirely obvious to follow), the victim's serious injury or death would be virtually certain to follow the defendant's act of shooting them in the ordinary course of event.

### *10.1.1. Direct and Oblique Intent*

"Direct" intention refers to the more familiar notion of subjective *mens rea*, although the term remains notoriously undefined in UK law. Padfield writes that intention is 'used in relation to consequences: thus, a person may be said to intend the consequences of his actions if he wants them to happen.'[2] Monaghan explains direct intent as 'one's aim or

[2] Nicola Padfield, *Criminal Law* (10th ed. Oxford University Press 2016), 47.

purpose';[3] Cross identifies direct intention where the defendant 'aims, desires, or makes the decision to bring about a particular consequences which is prohibited.'[4] The definition of direct intent perhaps carrying the most authoritative pedigree is provided in *R v Mohan*[5] as a 'decision to bring about, in so far as it lies within the accused's power, the commission of the offence…, no matter whether the accused desired that consequence of his act or not.'[6] However, the "golden rule" with regards to jury directions is that the judge ought to 'avoid any elaboration or paraphrase of what is meant by intent, and leave it to the jury's good sense to decide whether [the defendant] acted with the necessary intent.'[7]

In those cases where the judge must give direction to the jury on the meaning of intention, however, the definition of oblique intention from *R v Woollin*[8] is provided, which states that the jury 'are not entitled to infer the necessary intention, unless they feel sure that [the prohibited consequence] was a virtual certainty (barring some unforeseen intervention) as a result of the defendant's actions and that the defendant appreciated that such was the case.'[9] The first limb of this test provides an objective definition of intention, and it is this that similarly provides the objective definition adopted in the present thesis. The second limb of the test in *Woollin* is not directly adopted, however, as it returns to an inquiry into the defendant's subjective state of mind by asking whether they in fact appreciated that the prohibited consequences were virtually certain to result from their actions.

The hybrid test of intention proposed in this thesis first adopts the objective definition from *Woollin*, defining intention as existing when particular consequences are the virtually certain result (barring any unforeseen intervention) of a given action. Second,

[3] Nicola Monaghan, *Criminal Law Directions* (6th ed. Oxford University Press 2020), 58; see similarly Law Commission, *Legislating the Criminal Code: Offences Against the Person and General Principles* (Law Com No 218, 1989), 8.
[4] Noel Cross, *Criminal Law and Criminal Justice: An Introduction* (SAGE Publications 2010), 33.
[5] *R v Mohan* [1976] QB 1.
[6] *Ibid*., 11; approved in *R v Pearman* (1984) 80 Cr App R 259; similar statements precursor in *Cunliffe v Goodman* [1950] 2 KB 237, 253; approved in *Hyam v Director of Public Prosecutions* [1975] AC 55, 74.
[7] *R v Moloney* [1985] 1 AC 905, 926.
[8] *R v Woollin* [1999] 1 AC 82.
[9] *Ibid*., 96; approving *R v Nedrick* [1986] 1 WLR 1025, 1028.

the hybrid approach asks whether it is reasonable to expect that *anybody* sharing the defendant's relevant circumstances would have appreciated that virtual certainty. Phrased differently from a negative perspective, the second limb of the hybrid test asks whether there are any relevant circumstances of the defendant which would reasonably be regarded as *negating* any person's (*i.e.,* the "reasonable man's") capacities for volition (*i.e.*, reasons responsiveness and ordinary self-control) or for appreciating the virtually certain consequences of their actions. This is crucially different to the entirely subjective approach to the second limb in *Woollin*, which asks whether the defendant *actually* appreciated the virtual certainty of consequences flowing from their actions. Under the hybrid approach, the second limb of the test is comparing the defendant's conduct to that reasonably expected from anybody else in society, albeit allowing for the hypothetical comparator to share relevant circumstances with the defendant.

It ought to be noted that, in most cases, a subjective intention to act can never be categorically proven without some form of confession or irrefutably probative evidence (such as a detailed written-out plan). Rather, intention is normally inferred from a defendant's actions; 'a punch on the nose will normally (though not always) be brought about by an intention to punch someone on the nose, so that outcome and intention are directly linked.'[10] For this reason, the proposed shift from an entirely subjective to predominantly objective conception of *mens rea* (albeit applied to subjective circumstances) is not as radical as it may first appear. Indeed, it is arguable that the significant majority of contested criminal trials will require the jury to infer subjective states of mind from evidence of the defendant's actions. Moreover, approaching intention through objective definition has strong support in legal literature, not least from the seminal jurist Glanville Williams who notes *inter alia* the acceptance of oblique intent in legal systems around the world.[11]

There are two key reasons why, notwithstanding the proposed hybrid approach, it will remain relevant and desirable to continue to adduce evidence concerning a defendant's subjective state of mind. First, on the rare occasions that a subjective state of mind can be

[10] Alan Norrie, *Crime, Reason and History* (2nd ed. Butterworths 2001), 47.
[11] Glanville Williams, 'Oblique intention' (1987) 46(3) *Cambridge Law Journal* 417, 421 – 422.

proved conclusively – for example, with a taped confession – it stands to reason that the objective definition of that relevant state of mind would be satisfied. An objective definition is inherently easier to satisfy, for which reason the hybrid approach must continue to consider the defendant's relevant subjective circumstances when applying the second limb of the test so as to return some restraint to its satisfaction. Thus, whilst proof of any subjective state of mind is no longer a *prerequisite* for legal responsibility under the present thesis, such proof would nonetheless necessarily satisfy the hybrid objective / subjective approach. Second, and relatedly, proof of subjective states of mind – including certain knowledge or beliefs – can continue to have evidential value in satisfying the hybrid tests for *mens rea*. Again, the crucial point is that proof of entirely subjective states of mind is no longer a necessary prerequisite for responsibility, but can nevertheless have evidentiary value.

### *10.1.2. Testing Hybrid Intention in Jurisprudence*

#### *10.1.2.1. Director of Public Prosecutions v Smith*

The case of *Director of Public Prosecutions v Smith*[12] provides a suitable starting point where, contrary to the current conception of intention as a typically subjective concept, the House of Lords 'used an objective presumption to conclusively identify what the defendant's intention was', thereby adopting an entirely objective approach to intention.[13] As to the relevant facts of the case, the defendant was driving with stolen property in his car and was instructed to pull over by a police officer. Instead, the defendant accelerated away whilst the police officer clung onto the defendant's car; the defendant continued to drive away erratically, and the police officer was thrown from the defendant's car into traffic and died. The defendant asserted that he had not been aware that the police officer was holding onto the car and, therefore, the officer's death was an accident; and, in any event, the defendant claimed that he drove erratically only with the intention of escaping arrest and not with any intention to kill or cause grievous bodily harm, thereby refuting *mens rea* for the offence of murder.

[12] *Director of Public Prosecutions v Smith* [1961] AC 290.
[13] Monaghan (2020), 60 – 61.

During the judge's summation for the jury at trial, Donovan J. directed 'if you are satisfied… that [the defendant] must as a reasonable man have contemplated that grievous bodily harm was likely to result to that officer… and that such harm did happen and the officer died in consequence, then the accused is guilty of… murder.' This direction was confirmed by the House of Lords and the defendant's conviction for murder was ultimately upheld. If the defendant's submission that his intention was only to escape were accepted, an entirely subjective formulation of intention would necessarily acquit, as the defendant lacked the specific intention to kill or cause serious harm. Conversely, an entirely objective formulation (as approved in *Smith*) results in conviction, because it is plain to the hypothetical reasonable man that driving away with somebody clinging to a vehicle is highly liable to result in their serious injury or death. It is equally likely that the defendant would be convicted under the approach to oblique intention from *Woollin*.[14] In an incredibly similar and current case involving defendants who drove away with a police officer caught in a rope trailing behind their car, the trial judge was clear that the defendants would have been guilty of murder if the prosecution had established that they *knew* the officer was caught behind their vehicle.[15] Such knowledge alone would be sufficient in the circumstances to establish oblique intention, even if the prosecution could not prove direct intention.

Kaveny offers three possible interpretations of the circumstances in *Director of Public Prosecutions v Smith*, against which the hybrid formulation of *mens rea* may be tested:

> '(1) Smith drove erratically with the purpose of causing the police officer great bodily harm by throwing him off the car…
>
> (2) Smith drove erratically with the purpose of knocking the officer off the car, in order to facilitate his escape… however, he did not intend to harm the officer. Instead, that harm was a side-effect of his intentional act…

[14] See Glanville Williams, 'The *mens rea* for murder: Leave it alone' (1989) 105(Jul) *Law Quarterly Review* 387.
[15] Justice Edis, '*R v Long, Bowers, Cole and King* – Sentencing remarks' (31st July 2020), 2.

(3) Smith was simply driving away as fast as he could in order to escape the danger of being arrested. He did not intend to dislodge the officer from the car, let alone to harm him.'[16]

It is likely that Smith would be responsible in all three scenarios; in each instance, causing death or serious injury is virtually certain to result from driving erratically with somebody clinging onto the outside of the vehicle (objective definition). Moreover, there was nothing submitted in the defendant's circumstances to suggest that anybody else in the same circumstances would not reasonably be expected to appreciate that virtual certainty (objective / subjective test). Indeed, the only argument raised by Smith which could have refuted this finding was the submission that he had been unaware that the police officer was clinging to the car; however, the jury dismissed this as lacking credibility. Thus, the hybrid approach applied to *Director of Public Prosecutions v Smith* reaches the same findings on criminal responsibility as the House of Lords in that case, and as would likely be found today under oblique intention (but not following a purely subjective test).

Is this result correct and appropriate? For one, there exists an apparently strong (and, presumably, well-reasoned) judicial sentiment in favour of conviction in *Smith*; not only did the judge at first instance and the House of Lords find a conviction of murder to be appropriate but further, as Lord Goff highlights extra-judicially, 'a very distinguished and experienced group of judges *felt* that Smith *could* be held guilty of the crime of murder, even if he did not in fact intend to kill his victim or to cause him grievous bodily harm. Such a judicial reaction is not lightly to be disregarded.'[17] Lord Goff similarly proceeds to support conviction in *Smith*, albeit via a different justification. Moreover, there are compelling arguments for why the three interpretations of *Smith* ought to be treated the same. For example, from the perspective of likelihood of causing harm, all three scenarios are arguably alike insofar as there is an equal likelihood of harm from driving erratically with another person clung to the vehicle, *for whatever purpose or intent*. Similarly, the scenarios are alike from the perspective of foreseeability of harm;[18] whatever Smith's

[16] M. Cathleen Kaveny, 'Inferring intention from foresight' (2004) 120(Jan) *Law Quarterly Review* 81, 102.
[17] Lord Goff, 'The mental element in the crime of murder' (1987) 22(1) *Israel Law Review* 1, 9.
[18] Kaveny (2004), 102.

purposes, harm is equally foreseeable across the three scenarios once it is determined that Smith knew the officer was clinging to the car, as the jury so concluded.

Stannard provides four arguments for stigmatising the murderous 'ruthless risk-taker' against which the three interpretations of *Director of Public Prosecutions v Smith* might be compared. The first argument from *equivalence* argues, like oblique intention, that a certain degree of foresight of consequences may amount to a 'species of intention.'[19] As argued above, the foresight for potential harm across all three interpretations is equivalent once it is accepted that the defendant was aware that somebody was clinging to the car as they drove erratically. The second argument from *choice* submits that the 'willingness of ruthless risk-takers to endanger life renders them deserving to be bracketed alongside those who set out to take it.'[20] Again, regardless of Smith's purpose, in all three interpretations he displays a willingness to endanger life by driving erratically with the officer clung onto the car. The third argument is similar in suggesting that, once somebody decides to cause harm to another a 'crucial *moral threshold* is already crossed, and there is good reason to impose liability for whatever consequences may ensue.'[21] The fourth and final argument from *attitude* suggests that the ruthless risk-taker displays a 'culpable indifference to the value of human life' which justifies a similar treatment to murder if death does in fact follow from their actions.[22] Applying this argument, again, there is little distinction between the three interpretations of *Smith*. On balance, therefore, it may reasonably be concluded that the hybrid interpretation of intention would correctly and accurately convict the defendant in *Smith*.

[19] John E. Stannard, 'Murder and the ruthless risk-taker' (2008) 8(2) *Oxford University Commonwealth Law Journal* 137, 137; citing Andrew Ashworth, *Principles of Criminal Law* (5th ed. Oxford University Press 2006), 177; H. L. A. Hart, 'Intention and punishment' in Hart H. L. A. (ed.), *Punishment and Responsibility: Essays in the Philosophy of Law* (2nd ed. Oxford University Press 2008), 120 – 122; Andrew P. Simester, 'Moral certainty and the boundaries of intention' (1996) 16(3) *Oxford Journal of Legal Studies* 445.

[20] *Ibid*., 138; citing Finbarr McAuley and J. Paul McCutcheon, *Criminal Liability: A Grammar* (Sweet and Maxwell 2000), [300-4].

[21] *Ibid*; citing Ashworth (2006), 87; John Gardner, 'Rationality and the rule of law in offences against the person' (1994) 53(3) *Cambridge Law Journal* 502; Jeremy Horder, 'A critique of the correspondence principle in criminal law' (1995) (Oct) *Criminal Law Review* 759.

[22] *Ibid*; citing Barry Mitchell, 'Culpably indifferent murder' (1996) 25(1) *Anglo-American Law Review* 64; Antje Pedain, 'Intention and the terrorist example' (2003) (Sep) *Criminal Law Review* 579.

It is interesting to note that the approach to objective intention in *Director of Public Prosecutions v Smith* does bear a number of close similarities with the hybrid approach proposed in this thesis. Specifically, the House of Lords was clearly minded that the objective test for intention would be applied with certain circumstances subjective to the defendant in contemplation. As Viscount Kilmuir LC commented, 'once the accused's knowledge of the circumstances and the nature of his acts has been ascertained, the only thing that could rebut the presumption [of intention] would be proof of incapacity to form an intent, insanity, or diminished responsibility.'[23] Thus, the House of Lords were minded that subjective characteristics such as the defendant's knowledge of the circumstances would be relevant to the application of an objective test for intention. It is further pertinent to note that the House of Lords considered this objective approach to intention, applied in the light of certain relevant subjective circumstances, to have a robust pedigree in jurisprudence.[24] The hybrid approach to *mens rea* might therefore be regarded as returning to something approaching orthodoxy in the modern common law discussion of intention.

#### *10.1.2.2. Hyam v Director of Public Prosecutions*

Following the House of Lords adoption of an entirely objective test for intention in *Director of Public Prosecutions v Smith*, Parliament indicated its dissatisfaction with this approach through the Criminal Justice Act 1967, which provided that a jury was 'not bound in law to infer that [a defendant] intended or foresaw a result of his actions by reason only of its being a natural and probable consequence of those actions; but shall decide whether he did intend of foresee that result by reference to all the evidence, drawing such inferences as may appear proper in the circumstances.'[25] This clearly indicated a return to subjective intention, albeit allowing for the jury to infer the same

[23] *Smith* [1961], 331.
[24] *Ibid*., citing *R v Faulkner* (1877) 13 Cox CC 550, 561 – 562; *R v Lamely* (1911) 22 Cox CC 635, 636; *R v Philpot* (1912) 7 Cr App R 140, 141 – 144; *Director of Public Prosecutions v Beard* [1920] AC 479; 503 – 504; *R v Ward* [1956] 1 QB 351, 356; Oliver Wendel Holmes, *The Common Law* (Cosimo Inc 2009), 53 – 56.
[25] Criminal Justice Act 1967, s. 8.

from proof of objective intention as the "natural and probable consequences" of an action.[26]

The issue of defining intention arose again in *Hyam v Director of Public Prosecutions*.[27] The facts of this case concerned a defendant who poured petrol through the letterbox of their ex-lover's fiancée's house and ignited a fire, ultimately killing two children who were inside the property. It was established that the defendant knew that there were occupants inside the property and further appreciated the possibility of causing them harm – the defendant had first checked that her intended lover was *not* at the house. In the wake of both *Director of Public Prosecutions v Smith* and the Criminal Justice Act 1967, the relatively narrow question before the House of Lords in *Hyam* was whether or not the requisite *mens rea* of intention to kill could be 'established by proof beyond reasonable doubt that when doing the act which led to the death of another the accused knew that it was highly probable that that act would result in death or serious bodily harm?'[28] Thus, the question concerned what degree of subjective foresight of consequences was required in order to establish oblique intention.

The defendant was convicted for murder and the House of Lords upheld this conviction, but only by a slim majority of three judges to two, and with each member of the panel providing often contradictory judgments. The majority approved the direction of Acker J at first instance, that the requisite *mens rea* for murder could be found where somebody does an act knowing that it is 'highly probable that he will cause death or grievous bodily harm.'[29] For Lord Hailsham, the 'real impetus' for upholding the defendant's conviction was the 'fact that she had caused the fatal consequences volitionally rather than (with a particular degree of) foresight.'[30] He considered that the requisite *mens rea* for murder consisted of an intention to cause death, an intention to cause grievous bodily harm or, 'where the defendant knows that there is a serious risk that death or grievous bodily harm

[26] Janet Loveless, Mischa Allen and Caroline Derry, *Complete Criminal Law: Text, Cases, and Materials* (7th ed. Oxford University Press 2020), 99 – 100.
[27] *Hyam* [1975] AC 55
[28] *Ibid.*, 66.
[29] David Ormerod and Karl Laird, *Smith, Hogan, and Ormerod's Text, Cases, and Materials on Criminal Law* (13th ed. Oxford University Press 2020), 98.
[30] Beatrice Krebs, 'Oblique intent, foresight and authorisation' (2018) 7(2) *UCL Journal of Law and Jurisprudence* 1, 8.

will ensue from his acts, and commits those acts deliberately and without lawful excuse, the intention to expose a potential victim to that risk as the result of those acts.'[31] The facts of *Hyam* clearly fell within the Lord Hailsham's third category. What 'made [the defendant's] mind guilty to a sufficient degree to warrant a murder conviction was not her appreciation of the risk as such, but the fact that she went on to *embrace* that risk… Mrs Hyam was not just taking risks she ought not to have taken; the creation of danger was central to her goal of teaching her rival a lesson.'[32]

Lord Cross and Viscount Dilhorn each relied largely upon concepts of oblique intention in reasoning to uphold the defendant's conviction. Where the House of Lords found agreement is in defining oblique intention using phrases such as 'foresight of probability', 'foresight of high probability' and 'foresight of a serious risk.'[33] These are notably lower standards than the current conception of oblique intention as virtual certainty; moreover, these descriptions deploy terminology more closely associated with recklessness than intention, something which became a particular focal point for criticism after the judgment.[34] The two dissenting Lords Diplock and Kilbrandon were more concerned to restrict the *mens rea* of murder only to an intention to kill or cause death.[35]

It is submitted that the defendant would also have been criminally responsible for murder under the hybrid objective / subjective approach to intention. First, it is readily appreciable that setting a house ablaze with occupants inside is virtually certain to result in death or serious injury, and there are further circumstances of the case which point towards this conclusion. The defendant set the fire at night when anybody in the house would likely be asleep; the defendant also checked to ensure that her intended lover was not at the property because she did not want to cause him harm, thus clearly recognising the risk that somebody may be injured by her actions. The defendant took steps to avoid making noise and waking anybody whilst at the property, and she made no attempt to

[31] *Hyam* [1975], 79.
[32] Krebs (2018), 8 – 9; see also Kaveny (2004), 98 – 99.
[33] Gerard Coffey, 'Codifying the meaning of "intention" in the criminal law' (2009) 73(5) *Journal of Criminal Law* 394, 397 – 398.
[34] For example, see A. K. W. Halpin, 'Intended consequences and unintentional fallacies' (1987) 7(1) *Oxford Journal of Legal Studies* 104.
[35] Krebs (2018), 9 – 10.

alert anybody after setting the blaze. Finally, whilst not explicitly mentioned, the judgement implies that the tribunal of fact had found that the defendant knew that her rival was at least present in the house when she set the fire, even if she did not appreciate that her rival's children were present also.[36] Of course, nobody could be absolutely certain of consequences in a scientific sense;[37] but, in the ordinary course of events and barring any additional intervention, causing death or serious by lighting ablaze a house with occupants inside can be appreciated as a virtual certainty. Fire may not only kill through burning, but fire may block escape, and smoke can incapacitate or, in the case of people already sleeping, kill before they are ever aware that any danger exists.

Second, there are no circumstances of the defendant presented which suggest that it would not be reasonable to expect anybody else in the same circumstances to appreciate the virtual certainty of consequences in that case. It was argued that the defendant was driven by anger and jealousy, but the force of these emotions alone is never *reasonably* permitted to excuse any defendant of responsibility for their actions. Furthermore, as Pedain argues, the fact that the defendant claims to have only intended to frighten their rival and did not want to cause any harm is immaterial. He describes two senses of not wanting something to occur; one sense in which a person actively wants something *not* to happen, and a second sense in which a person is indifferent to something occurring, *i.e.*, they did not *actively* want it even though it would occur nonetheless. The defendant in *Hyam* did not want to cause harm in the second sense; however, as Pedain writes, 'that in itself is insufficient for her to disassociate herself from this consequence of her intentional conduct – in fact it does nothing towards it.'[38] Thus, the hybrid approach to intention would again reach the same conclusions as the House of Lords in *Hyam* and correctly uphold the defendant's conviction for murder.

[36] *Hyam* [1975], 63.
[37] Pedain (2003), 587.
[38] *Ibid*.

#### *10.1.2.3. R v Mohan and R v Belfon*

The cases of *R v Mohan*[39] and *R v Belfon*[40] are factually unrelated, yet warrant consideration together. Both cases were decided by the Court of Appeal under the binding authority of *Hyam*, whilst neither case concerned the offence of murder. Indeed, it was for this latter reason that the Court of Appeal distinguished both *Mohan* and *Belfon* from *Hyam*, thereby escaping the binding House of Lords authority and attempting to 'limit the application of the wide definition of *Hyam*.'[41] It is also notable that the defendants in both *Mohan* and *Belfon* were convicted on multiple charges and, whilst the Court of Appeal was bound to quash some of these charges due to insufficient directions given by the trial judge, the appellate Court considered it to be nonetheless fortunate in the interests of justice that each defendant would continue to serve their deserved sentences. Thus, the Court was plainly of the view that conviction and punishment were appropriate in each case, even if particular charges needed to be overturned on technical grounds.

In *R v Mohan*, the defendant slowed his car in response to a police officer's stop signal but then, within ten metres or so, rapidly accelerated the car towards the officer, forcing him to jump out of the way. The defendant was convicted of multiple charges, from which the point of contention concerned the offence of attempting to cause bodily harm by wanton driving.[42] The particular contention – regarding which the trial judge incorrectly directed that the offence could be committed recklessly whereas, as the Court of Appeal concluded, only intention would suffice[43] – is not pertinent to the present discussion. However, accepting the Court's premise that the offence requires intention and not recklessness, the facts can be assessed under the proposed hybrid approach to intention. First, it is readily arguable that bodily injury is virtually certain to result from accelerating a vehicle towards somebody suddenly and from a short distance away. Second, there is nothing presented in the defendant's subjective circumstances which would reasonably be accepted as diminishing anybody's appreciation of this virtual certainty. Thus, in concurrence with the original verdict and, arguably, the sentiment of the Court of Appeal,

[39] *Mohan* [1976] QB 1
[40] *R v Belfon* (1976) 63 Cr App R 59.
[41] Padfield (2016), 49.
[42] Offences Against the Person Act 1861, s. 35.
[43] See further Kenneth J. Arenson, 'The pitfalls in the law of attempt: A new perspective' (2005) 69(2) *Journal of Criminal Law* 146.

the defendant in *Mohan* would have been found similarly responsible under the hybrid approach to intention.

In *R v Belfon*, the defendant and another set upon a group of people exiting a public house, in particular slashing at the victims with an open razor and causing severe injuries to the head and face. The defendant was, again, convicted on a number of charges, from which the point of contention concerned the offence of wounding with intent to cause grievous bodily harm.[44] The particular contention is, again, not pertinent to the present discussion, but the Court of Appeal acquitted the defendant of this charge because the trial judge had indirectly directed that the offence could be committed with recklessness whereas only intention would suffice. However, once more it is clear that the hybrid approach to intention would uphold the original conviction and the underlying sentiment of the Court. Grievous bodily harm – which is to say "serious" harm – is (first) virtually certain to result from slashing at somebody's face with an open razor, and (second) there are no circumstances presented which would reasonably diminish anybody's capacity to appreciate this virtual certainty.

The value of considering these cases together is that the Court of Appeal was clearly concerned with limiting what it considered to be an overly broad interpretation of intention from *Hyam*. The definition provided in *Hyam* drew intention too close to recklessness by allowing the former to be concluded from foresight of highly probable risk. This helps inform the hybrid approach to intention by adopting a narrower objective definition of "virtual certainty", clearly distinguishing intention from recklessness. That notwithstanding, the application of the hybrid approach to the facts of *Mohan* and *Belfon* continues to demonstrate how this approach would have arguably reached the appropriate finding of responsibility in both cases. In this regard, it is again highlighted that the Court of Appeal overturned convictions on particular charges in each case due to incorrect directions provided by the trial judge; but the Court otherwise remained satisfied that each defendant would continue to be punished as justice demanded.

[44] Offences Against the Person Act 1861, s. 18.

#### *10.1.2.4. R v Moloney*

The next case of *R v Moloney*[45] arose from undisputedly peculiar facts. The defendant and his stepfather had been drinking into the night following a family celebration. According to the defendant's account, they had discussed the defendant's desire to leave the army and, over the course of a protracted and increasingly drunken conversation, the topic turned to their respective prowess with a gun. The challenge was purportedly thrown down by the defendant's stepfather that he could outshoot, outload and outdraw the defendant, and so the defendant retrieved two shotguns so that the challenge could be tested. In the defendant's statement to the court, recorded from interview, he recalled:

> '[My stepfather] opened his gun and started to remove his snap caps. I opened my gun and removed two empty cartridges which I used as snap caps as I don't have any. I inserted the cartridge in the right-hand barrel, closed the gun, took off the safety catch and pulled the trigger of the left-hand barrel, and told him he'd lost. By this time I don't think he'd even cleared his barrel of the snap caps. He looked at me and said: "I didn't think you'd got the guts, but if you have pull the trigger." I didn't aim the gun. I just pulled the trigger and he was dead. I then went and called the police and told the operator I had just murdered my father, and that's the story.'[46]

The defendant was convicted of murder following the wide interpretation of intention that was the current law under *Hyam*. Providing the leading judgment in *Moloney*, Lord Bridge first stated unequivocally that foresight of consequences is not at all the same thing as intention and belongs, 'not to the substantive law, but to the law of evidence.'[47] Whilst Lord Bridge continues to repeat the golden rule that directions on intention ought to be avoided wherever possible, however, the direction that he proceeds to offer appears to return right back to imputing foresight of risk with intention. He directs that the jury must answer two questions:

---

[45] *Moloney* [1985] 1 AC 905.
[46] *Ibid*., 916.
[47] *Ibid*, 928; see further Halpin (1987), 109 – 110; A. D. Chantry, '*R v Moloney* and the mental element in murder' (1985) 7(2) *Liverpool Law Review* 168, 177.

> ‘First, was death or really serious injury in a murder case (or whatever relevant consequences must be proved to have been intended in any other case) a natural consequence of the defendant’s voluntary act? Secondly, did the defendant foresee that consequence as being a natural consequence of his act?’[48]

In the event, the defendant’s conviction for murder was considered to be unsafe due to misdirection by the trial judge. However, the House of Lords expressed their contentment that reducing the conviction to manslaughter achieved justice in the case, as there could be no doubt that the defendant had unlawfully killed his stepfather following the wholly reckless, if not unlawful, behaviour of drunkenly playing with guns.

Applying the hybrid approach to intention to the facts of *Moloney*, there is a strong possibility that the defendant would similarly not be found responsible for murder but would undoubtedly be responsible for manslaughter. Starting with the *prima facie* position, it is first arguable that causing death is virtually certain to following from firing a shotgun towards another person’s head at close range. Regarding the second limb, however, the defendant’s formal defence amounted to a bare denial of *mens rea* in which he was asserting that he never deliberately aimed the gun, and even less so towards his stepfather.[49] If this assertion is accepted as credible (discounting for the moment the defendant’s intoxication), the relevant question under the second limb asks whether it is reasonable to expect anybody who had not deliberately aimed a gun at somebody else to appreciate the virtual certainty of killing another when the gun was fired. It is submitted that the natural answer to this question would be in the negative; whilst firing a gun without properly aiming undoubtedly amounts to a higher degree of recklessness, it is far from virtually certain that somebody will be injured and, therefore, it would be unreasonable to expect anybody in the defendant’s circumstances to appreciate the virtual certainty of killing another person by firing a gun that had not been aimed.

[48] *Moloney* [1985], 929.
[49] *Ibid.*, 916 – 917.

The defendant's second defence in *Moloney* was a denial of specific intention on account of intoxication; the fact that this defence was not adequately put to the jury formed part of the basis for the defendant's successful appeal and substitution for a verdict of manslaughter. This would similarly be a relevant subjective circumstance of the defendant for consideration in the second limb of the hybrid approach to *mens rea*. Thus, the question becomes whether it is reasonable to expect anybody sharing the defendant's degree of intoxication to appreciate the virtual certainty of killing another by firing a gun aimed towards them. Ultimately, this would be a matter for which the jury would have to draw the appropriate line; however, there are three reasons why a high degree of intoxication would be reasoned to undermine a specific intention for murder in this case.

First, repeating the arguments for the defendant's first defence, intoxication may be submitted to support the defendant's claim that they were unaware of the gun being aimed towards the victim, in which case the same conclusions follow. Furthermore, a sufficient degree of intoxication may undermine the presumption of volition such that, second, the defendant is no longer regarded as being appropriately reasons responsive or, third, the defendant no longer possessed the requisite ordinary degree of self-control. Therefore, the hybrid approach to intention would again likely return a similar substantive result in *Moloney* of reducing the defendant's conviction from murder to manslaughter. In addition, this example demonstrates how intoxication could diminish the *specific* intent required for certain crimes, just as is currently accepted in UK law.[50] The application of the intoxication defence is considered in greater detail in section 11.3.3 of this thesis, below.

#### *10.1.2.5. R v Hancock and Shankland*

It is notable that the decision in *Moloney* quickly came under attack, not least because the direction provided by Lord Bridge once again appeared to define intention closely to recklessness and, furthermore, because it remained ambiguous regarding what amounted

[50] See Arlie Loughnan and Nicola Wake, 'Of blurred boundaries and prior fault: Insanity, automatism and intoxication' in Reed A., Bohlander M., Wake N. and Smith E. (eds.), *General Defences in Criminal Law: Domestic and Comparative Perspectives* (Ashgate Publishing 2014), 115 – 116; citing *Director of Public Prosecutions v Majewski* [1977] AC 443; *Beard* [1920].

to the "natural consequences" of an action.[51] Only one year later, the House of Lords took the opportunity to provide further clarification in *R v Hancock and Shankland*.[52] Against the backdrop of the 1980s miners' strikes, the defendants in this case were striking miners who pushed a block of concrete and a concrete post from a bridge overlooking a road, along which another miner was being driven to work by taxi. The projectiles hit the windscreen of the taxi and killed the driver. At trial, the defendants asserted that they only intended for the projectiles to land in the road and frighten the miner, who was going to work contrary to the general strike; therefore, they denied that they had any intention to kill or cause serious harm to anybody. The defendants were subsequently convicted of murder following the directions on intention laid down in *Moloney* the previous year.

The Court of Appeal allowed the defendant's appeal on the grounds that, *inter alia,* the direction on intention provided in *Moloney* was ambiguous and misleading, and this point was subsequently upheld by the House of Lords. As Pigott explains, in his judgment in *Moloney* Lord Bridge uses the word "natural" 'in a special sense which it certainly does not convey without explanation.'[53] In fact, Lord Bridge provided this explanation earlier in his judgment when he remarked that "natural" 'conveys the idea than in the ordinary course of events a certain act will lead to a certain consequence unless something unexpected supervenes to prevent it'[54] and that the 'probability of the consequence taken to have been foreseen must be little short of overwhelming before it will suffice to establish the necessary intent.'[55] This was not, however, contained within the guidance directions that Lord Bridge ultimately laid down in *Moloney* which, consequently, the House of Lords considered to be unsafe one year later in in *Hancock and Shankland*. In particular, Lord Scarman (providing the leading judgment) considered that any judicial guidance relating to the definition of intention needed to make reference to probability, *i.e.,* foresight of probable consequences.[56] Specifically, guidelines 'require an explanation that the greater the probability of a consequence the more likely it is that the consequence

[51] Coffey (2009), 399 – 400; Krebs (2018), 11 – 12; see also Nicola Lacey, 'A clear concept of intention: Elusive or illusory?' (1993) 56(5) *Modern Law Review* 621.
[52] *R v Hancock and Shankland* [1986] AC 455.
[53] Maggy Pigott, 'Murder – intention' (1986) (Mar) *Criminal Law Review* 180, 182.
[54] *Moloney* [1985], 929.
[55] *Ibid*., 925.
[56] *Hancock and Shankland* [1986], 472 – 475.

was foreseen and that if that consequence was foreseen the greater the probability is that that consequence was also intended.'[57] The House of Lords therefore commuted the defendant's conviction to manslaughter.

Here, again, it is submitted that the hybrid approach to intention would similarly have precluded a conviction for murder in favour of a conviction for manslaughter. This conclusion can potentially be reached under just the first limb of the test, asking whether the death of the taxi driver was a virtually certain result of throwing the concrete projectiles from the bridge. Undoubtedly, the actions were highly reckless and warranted a manslaughter conviction; however, it is quite arguable that the objective definition of intention as a virtual certainty would not be satisfied in this case. The taxi was a moving target, and the windscreen in particular was a small moving target, so the defendants would have needed very high skills of aiming and timing or a good degree of luck in order to throw the projectiles with a *virtual certainty* of killing the taxi driver, no less so considering that those concrete projectiles were heavy and cumbersome.

Nonetheless, if it is concluded that the victim's death was a virtual certainty of throwing concrete projectiles from a bridge, the inquiry moves to the second limb of the hybrid test. The only relevant circumstances provided by the defendants was the assertion that they had only meant to frighten the miner going to work and, therefore, had not aimed the projectiles at the taxi but into the road. In this sense, evidence of the defendants' state of mind is relevant to (albeit not determinative of) responsibility. If the jury is satisfied that the defendant's assertions are credible, then the question under the second limb of the hybrid test becomes, *is it reasonable to expect anybody who was aiming a projectile at the road in front of a moving vehicle to appreciate that death or serious injury was virtually certain to result from their actions?* This, it is submitted, could reasonably be answered either way, and it is upon this dividing line that the jury serves its key social function in the criminal law by determining the precise boundaries of responsibility.[58]

[57] *Ibid*., 473; see further Maggy Pigott, 'Intention – murder – model direction laid down in Moloney unsafe and misleading' (1986) (Jun) *Criminal Law Review* 400, 401.
[58] For example, see Kaveny (2004), 96.

The jury might conclude in the affirmative – even if the defendants only aimed their projectile in front of the moving vehicle, those actions (*i.e.,* pushing falling objects into traffic on the road) were so inherently dangerous that the mere fact of intending to aim for the road instead of directly at the vehicle would not excuse any other reasonable person from failing to appreciate the virtual certainty of death or serious injury following from their actions. Alternatively, the jury might conclude in the negative – the defendants aimed away from the vehicle, and the vehicle was moving and, therefore, a difficult target in any event. Thus, if it is accepted that the defendants were aiming specifically not to hit the vehicle, it might be argued that any other person similarly aiming not to hit a difficult moving target would also not reasonably be expected to appreciate that causing death was virtually certain to result from their actions. This is an undoubtedly difficult question to answer and, on balance, it is proposed that the latter negative answer is the more natural conclusion. Anybody aiming *not* to hit something – and, no less, aiming *not* to hit a difficult moving target with a heavy object – might rightly be surprised if they did indeed hit that target; they would have missed their aim – *i.e.,* the empty space – which was a considerably larger target than the smaller moving vehicle. It is submitted, therefore, that the reasonable man aiming not to hit the difficult moving target would not necessarily appreciate that death was virtually certain to ultimately follow from their action of throwing the projectile.

#### *10.1.2.6. R v Nedrick*

The facts of *R v Nedrick*[59] are virtually identical to those in *Hyam*. The defendant poured paraffin through the letterbox of another's house at night, purportedly to frighten them but with no explicit intention to cause serious injury or death. However, the defendant did hold a grudge against his intended victim, had previously threatened to "burn her out", and set the house ablaze during the night and with no warning to the sleeping occupants. In the event, a child sleeping in the house died and the defendant was convicted of murder. However, noting the date of the case, this conviction was delivered prior to the rulings in *Moloney* and *Hancock and Shankland* and, therefore, following defective guidelines. The defendant's conviction for murder was therefore commuted to manslaughter, but he was

[59] *Nedrick* [1986] 1 WLR 1025

nonetheless sentenced to 15 years imprisonment to reflect the seriousness of the offence. In this respect, it is notable that the defendant's only formal defence was that he had 'neither started the fire nor made any admissions to that effect,' which was entirely rejected by the jury.[60]

Mindful of those subsequent rulings, the Court of Appeal in *Nedrick* provided what has broadly become the currently accepted description of oblique intention, with some further minor clarification provided later in *R v Woollin*. First, the Court of Appeal confirmed that, for the offence of murder, the prosecution had to prove that the defendant intended to kill or cause serious injury; any finding of oblique intention therefore provided evidence upon which a jury could infer intention, but did not oblige them to do so – oblique intention provides evidence, but not proof, of direct intention. Second, where it is necessary to direct the jury, the pertinent questions are (1) *'how probable was the consequence which resulted from the defendant's voluntary act?*' and (2) *'did [the defendant] foresee that consequence?'*[61] Third, where the defendant did not appreciate that death or serious harm would result from his actions or thought that such a risk was only slight, it may be easier to infer that he did not intend to bring about the prohibited result. However,

> '[I]f the jury are satisfied that at the material time the defendant recognised that death or serious harm would be *virtually certain (barring some unforeseen intervention) to result from his voluntary act*, then that is a fact from which they may find it easy to infer that he intended to kill or do serious bodily harm.'[62]

It is, again, notable that the House of Lords maintained a relatively high sentence to reflect the severity of the defendant's behaviour, notwithstanding that the conviction was reduced to manslaughter on technical grounds. It is proposed that the hybrid objective / subjective approach to intention would maintain a conviction for murder, applying the

[60] *Ibid*., 1026.
[61] *Ibid*., 1028.
[62] Lynne Knapman, 'Murder – dangerous act – foresight of death or serious bodily harm' (1986) (Nov) *Criminal Law Review* 742, 742 – 743; citing *Nedrick* [1986], 1028.

same reasoning from *Hyam*, above. In brief, the defendant was aware that people were in the house, and committed the act in the early hours when it was most likely that people would be asleep; and they set the fire without giving any prior or subsequent warning. It may readily be argued that causing death or serious injury is virtually certain to result from setting ablaze a property with people sleeping inside; the fumes of the fire are liable to render people unable to wake up, whilst the positioning of the fire at the door of the property blocked the principal means of exit for anybody who did awaken. Meanwhile, there are no relevant circumstances presented which would suggest that it is unreasonable to expect anybody in the same circumstances to appreciate the virtual certainty of death or serious injury following their actions.

#### *10.1.2.7. R v Woollin*

In the final case to consider on intention, *R v Woollin*,[63] the defendant lost their temper at their crying three-month-old son and shook the infant before throwing him onto the hard floor. The child suffered from a fractured skull and subsequently died, and the defendant was tried and convicted for murder. Notably, the prosecution did not contend that the defendant had the direct intention to kill or cause serious injury, which the defendant denied in any event. Rather, the prosecution case was that, following *Nedrick*, the child's death or serious injury was virtually certain to follow from the defendant's actions of throwing him on the floor, and that the defendant had appreciated as much to be the case, thus resting the prosecution case solely on oblique intention. It is further notable that, as well as denying the requisite *mens rea* of intention, the defendant forwarded a positive defence of provocation which the jury ultimately rejected. The key issue on appeal was that whilst the trial judge had initially applied the direction on oblique intention from *Nedrick*, he then proceeded to introduce the language of recklessness in requiring the jury to be satisfied that the defendant 'must have realised and appreciated when he threw that child that there was a *substantial risk* that he would cause serious injury to it.'[64]

[63] *R v Woollin* [1999] 1 AC 82.
[64] *Ibid.*, 88.

The House of Lords considered that this misdirection rendered too unsafe the conviction for murder, reducing the conviction to manslaughter. The Court clarified what is, today, the accepted formulation of oblique intention, which exists when the prohibited outcome 'was a virtual certainty (barring some unforeseen intervention) as a result of the defendant's actions and that the defendant appreciated that such was the case.'[65] It is notable that the Court of Appeal did not consider the misdirection to be so grave as to render unsafe the defendant's conviction for murder, opining that the use of the phrase "virtual certainty" may not be necessary in every case.[66] It is equally notable that the Court of Appeal had on previous occasions similarly declined to differentiate between virtual or moral certainty and a very high degree of probability,[67] tacitly suggesting that the Court of Appeal was correct to uphold the conviction for murder in *Woollin*. Nevertheless, the House of Lords disagreed, stating that 'by using the phrase "substantial *risk*" the judge blurred the line between intention and recklessness, and hence between murder and manslaughter.'[68] Instead, the Court commented that the aforementioned definition consisting of virtual certainty ought to be given on the rare occasions when the jury requires a direction on intention.[69]

It is clear at this stage to appreciate the heritage of the proposed hybrid formulation of intention. First, the objective definition lifts directly from the direction in *Woollin*, defining intention as being where particular consequences are the virtually certain result (barring some unforeseen intervention) of a given action. However, the *Woollin* formulation proceeds to require proof that the defendant was *actually subjectively aware* of that virtual certainty, whereas such an entirely subjective approach has been rejected in this thesis. Instead, the hybrid objective / subjective test asks, second, whether or not it is reasonable to expect that anybody in the same relevant circumstances as the defendant would appreciate the virtually certain consequences of their actions. In the event, it is submitted that the hybrid approach would agree with the outcome at first instance and before the Court of Appeal in finding the defendant responsible for murder. To begin,

[65] *Ibid*., 96.
[66] *R v Woollin* (1997) 1 Cr App R 97, 105 – 107.
[67] *R v Walker and Hayles* (1990) 90 Cr App R 226, 232.
[68] *Woollin* [1999], 95.
[69] See further John C. Smith, 'Case commentary: *R v Woollin*' (1998) (Dec) *Criminal Law Review* 890.

there can be no doubt that the objective definition of intention is satisfied; throwing a three-month-old infant against a hard floor is virtually certain to result in death or really serious injury.

Proceeding to the second limb, relevant circumstances supporting the prosecution's case included the fact that the defendant had given multiple significantly differing accounts of how his son became injured to different people – doctors, paediatricians, and across numerous police interviews – calling into question his general credibility. Relevant circumstances supporting the defendant's case included, first, his bare denial of any intention (or foresight) in causing serious injury to the baby at the moment of his action due to being in a rage and, second, that he had lost control, applying the partial defence of provocation.[70] The second legal defence of provocation was rejected by the jury on the facts. Furthermore, it is plain that the new statutory defence of "loss of control", which replaces provocation, would be equally inapplicable in the present case as the actions of an infant could never amount to a "qualifying trigger" within the meaning of the relevant legislation.[71] If credible, the defendant's initial bare denial of intention due to rage would render the second limb of the hybrid test as asking, *is it reasonable to expect anybody in a fit of rage to nonetheless appreciate that throwing an infant to the ground is virtually certain to result in serious injury or death?*

It is submitted that the answer to this question must be in the affirmative. Society reasonably expects people to exert a degree of control over their behaviour even when experiencing intense anger, frustration or rage. And, indeed, such self-control is patently within the capacities of ordinary adults; most cases of such intense anger, frustration or rage may result in rash decisions, ill-considered actions and unwise behaviour *etc.*, but it is considerably less usual for this to evolve into patently dangerous and harmful conduct to others, and even less so towards infants and children. If this were not the case, the criminal courts might be considerably more inundated with cases. Moreover, there is good reason why the law should not want feelings of rage alone to amount to a legal defence;

[70] *Woollin* (1997), 100 – 101.
[71] Coroners and Justice Act 2009, ss. 54 – 56.

how easy might it be for every defendant to simply claim that they did not intend their actions, but they were in a rage.

Indeed, the law has addressed precisely this problem by strictly defining legal defences, and the question of how defences fit within the present thesis is considered in chapter eleven, below. In *Woollin*, however, defences that concern a loss of self-control over actions as such insanity, diminished responsibility, and the prior defence of provocation, were each inapplicable or, in the latter case of provocation, was found to be not proven by the jury. Therefore, is it *reasonable* to expect that anybody in a fit of rage would nonetheless appreciate that throwing an infant to the ground is virtually certain to result in serious injury or death? Absolutely; that is an entirely reasonable level of conduct for the criminal law to expect and require, and is entirely commensurate with the capacity for self-control presumed for all adults.

### *10.1.3. Final Comments on Intention*

As the discussion on intention aims to demonstrate, it is submitted that the hybrid objective / subjective approach to *mens rea* could reasonably have delivered the desirable result in each case, without the various to-and-froing between different definitions of intention within the courts. Aside from defining precisely what degree of foresight of consequences is necessary to permit a finding of intention, a second question has plagued the courts over whether or not oblique intention, so defined, amounts to the equivalent of intention or merely provides evidence from which intention may be deduced. Clearly, the hybrid approach adopted in this thesis equates intention with its hybrid objective / subjective formulation; indeed, actual subjective or direct intention is no longer a *necessary* component of *mens rea* under the hybrid approach, albeit the same would remain evidentially valuable and likely *sufficient* for satisfying the hybrid conception of intention.

Duff provides an interesting perspective on the question of whether or not legal intention should include foresight of 'moral certainties,' unlike ordinary or direct intention; he suggests that the issue 'reflects an underlying tension between two conceptions of agency

and responsibility.'[72] On the one hand, a more consequentialist approach focuses primarily on effects and places emphasis on knowledge and control; thus, 'I am paradigmatically responsible (the agent of) those effects which I foresee and over which I have effective control – for another's death if it is the foreseen and avoidable effect of what I do.'[73] On the other hand, a more deontological approach is focused primarily on actions and emphasises direct intention; 'the moral character of my actions depends crucially on the (direct) intentions – the quality of will – which they reveal.'[74] From this perspective, the various to-and-froing between different definitions of intention considered across decades of jurisprudence may be understood as reflecting different dominant approaches to the broader question of agency and responsibility.

The hybrid approach to *mens rea* supported in this thesis is undoubtedly more consequentialist than it is deontological. The objective definitions of each form of *mens rea* provide a descriptive, consequentialist account of when outcomes may be regarded as following actions with varying degrees of certainty or likelihood. Intention – *virtual certainty* – represents the greatest likelihood with which consequences will follow actions. However, the second limb inquiring whether or not it is reasonable to expect anybody in the defendant's circumstances to appreciate such virtual certainty goes further. This component relates relevant circumstances – objective circumstances, but also including the circumstances of what can credibly be established that the defendant subjectively knew, believed or foresaw *etc.* – to how those circumstance cause different behaviour. To speak of legal intention, therefore, is not to say what a defendant *actually* intended, but whether all the relevant circumstances at the time would indicate that any person should reasonably appreciate the certainty of consequences to follow from their actions. Duff expresses a similar point thus:

> '[W]e should rather understand intention as a matter of the relation between the agent's actions and her beliefs – which beliefs are relevant to, as providing the reasons for, her actions; and as a matter, not of what is going

[72] R. Anthony Duff, 'The obscure intentions of the House of Lords' (1986) (Dec) *Criminal Law Review* 771, 780.
[73] *Ibid*.
[74] *Ibid*.

on in her mind, but of the pattern which her actions (including what she says about what she is doing) instantiate.'[75]

The hybrid approach to intention (and *mens rea* generally) supported in this thesis broadly enacts Duff's approach; intention is no longer directly a question of what was within a defendant's subjective state of mind, but a question of whether the entire circumstances – including the defendant's pattern of actions and claimed subjective experiences – are such that anybody in the same circumstances would reasonably be expected to appreciate the virtually certain consequences of their actions. It is submitted that this shift in approach enables a hybrid objective / subjective definition of intention (and *mens rea* generally) to avoid the many pitfalls that have befallen the courts over decades of attempting to singularly define intention.

## *10.2. Recklessness*

Under the revised hybrid formulation, recklessness is defined objectively as occurring when *there is an unreasonable risk that the prohibited criminal outcome would result from the defendant's act*, and is assessed with the defendant's relevant subjective circumstances in mind by asking, *is it reasonable to expect anybody in the defendant's circumstances to appreciate that unreasonable risk*?

Whether or not a given risk is "unreasonable" will be determined by reference to the entire circumstances of the case; however, four considerations are likely to be relevant in most instances – the likelihood of that risk manifesting, the severity of resulting harm, the obviousness of a particular risk, and the utility in taking that risk. Thus, a risk consisting of a high probability of causing significant harm is patently unreasonable. Moreover, a risk consisting of a high probability of causing relatively minor harm might nonetheless be regarded as unreasonable – it is not generally regarded as reasonable conduct to cause others any harm whatsoever, however minimal, and even cutting another's hair may be regarded as an assault occasioning actual bodily harm, whilst assault can be committed

[75] *Ibid*.

intentionally or recklessly.[76] Therefore, a *high* probability of causing even minimal harm might rightly be regarded as unreasonable. And, *vice versa*, a risk consisting of a lower probability of causing relatively significant harm might equally be regarded as unreasonable; even though the probability of harm is lowered, the increased severity of harm means that it is less reasonable to take such risks.

The more obvious any risk of harm is, the more likely it is that that risk is unreasonable, whereas it is more "reasonable" – or, at least, more excusable – to take risks that are less obvious because, by definition, the less obvious any risk is, the less reasonable it is to expect anybody to be able to foresee that risk. Finally, many risks may be regarded as being more reasonable or acceptable when they also carry a higher degree of utility. For example, over-taking on a busy road might carry a rather high probability of causing significant harm, but could also carry considerable utility, for example, in the case of an ambulance rushing to an emergency. Thus, it is plain to see that the *reasonableness* of any particular risk will be a highly context-dependent question to be determined by the jury as a question of fact in any given case.

### *10.2.1. Subjective and Objective Recklessness*

Much like with the definition of intention, the courts have struggled for decades to settle upon a single accepted definition of recklessness, and have even applied both an objective and a subjective test to different offences for a period of time.[77] A subjective test currently prevails, with a person being regarded as acting recklessly with respect to '(i) a circumstance when he is aware of a risk that it exists or will exist; (ii) a result when he is aware of a risk that it will occur; and it is, in the circumstances known to him, unreasonable to take the risk.'[78] In brief, the objective definition of recklessness that existed for some time provided that a person was reckless with regards to a particular action or outcome if they acted to create an obvious risk of that circumstance occurring,

[76] *Director of Public Prosecutions v Smith* [2006] EWHC 94; Offences Against the Person Act 1861, s. 47.
[77] See further Jonathan Herring, *Criminal Law* (11th ed. Red Globe Press 2019), 70 – 73.
[78] *R v G* [2003] UKHL 50, [41].

whether he has 'not given any thought to the possibility of there being any such risk or has recognised that there was some risk involved and has none the less gone on to do it.'[79]

The House of Lords' affirmation of the subjective test notwithstanding, it is not entirely clear that this subjective approach always leads the way in practice before the courts. For one, the authority of *R v Parker*[80] – discussed further in the following section, below – continues to suggest that defendants may be found reckless towards risks that they did not subjectively foresee if they deliberately closed their mind towards the existence of a patently obvious risk, inviting an inherently objective assessment. Equally, it is well established under the current law that a defendant may not claim that they did not foresee an obvious risk simply because they were *voluntarily* intoxicated.[81] Rather, a voluntarily intoxicated defendant is deemed to have foreseen those risks which they would have otherwise foreseen had they been sober, again pointing towards a more objective assessment. This is particularly significant given the prevalent association between alcohol and criminal behaviour: for example, more than a million offences *per annum* in the UK are associated with alcohol intoxication, including between 39% and 54% of violent crimes across different parts of the country.[82]

Furthermore, within one month of the House of Lords' confirmation of the subjective test for recklessness in 2003, the Sexual Offences Act 2003 received Royal Assent with a new definition of rape including the absence of a reasonable belief in consent. As Lio explains, 'a failure to realise that the victim does not consent, cannot be justified because of lack of imagination, stupidity, or "honest mistakes"… [t]his is the very reason why Parliament was urged to amend the offence of rape in the 2003 Act.'[83] This again reintroduces

[79] *Commissioner of Police of the Metropolis v Caldwell* [1982] AC 341, 354.
[80] *R v Parker* [1977] 1 WLR 600.
[81] *Majewski* [1977]; *R v Bennett* [1995] Crim LR 877.
[82] Institute of Alcohol Studies, 'Crime and social impacts' (IAS 2019) <http://www.ias.org.uk/Alcohol-knowledge-centre/Crime-and-social-impacts.aspx#_ftn2> accessed 10 November 2020; citing Office for National Statistics, 'The nature of violent crime in England and Wales: year ending March 2018' (ONS 2018) <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/thenatureofviolentcrimeinenglandandwales/yearendingmarch2018> accessed 10 November 2020; Alcohol Health Alliance UK, 'Measuring up: The state of the nation' (Alcohol Health Alliance UK 2017) < https://ahauk.org/wp-content/uploads/2017/12/7119-AHA-10-year-anniversary-report.pdf> accessed 5 October 2022.
[83] Jennifer Lio, '*Cunningham* recklessness: The quintessence of the historic English criminal law?' (2018) 6 *North East Law Review* 71, 73.

inadvertent recklessness within the criminal law, only moments after the House of Lords had finally dispensed with objective recklessness and, 'at that, in the context of a rather more serious offence.'[84] Thus, despite the House of Lords pronouncing that the test for recklessness is subjective, it is nonetheless evident that objective approaches to recklessness remain throughout the criminal law.

The emergence of two different tests for recklessness might be best understood through an appreciation of their respective historical developments; for example, Cunningham describes two concurrent developments of the legal concept of recklessness.[85] She comments how recklessness was first used as a measure of blameworthiness in criminal law 'as an epithet in relation to the degree of negligence required to find a defendant guilty of manslaughter', from which arose the question of whether an objective test ought to be applied.[86] In the case of *R v Williamson*,[87] an untrained male midwife mistook a prolapsed uterus for part of the placenta and inadvertently killed his patient in attempting to sever the bodily material. The defendant was acquitted of manslaughter because there was no evidence that he had been inattentive, from which point the law started to develop a more objective test requiring that the defendant 'display a certain degree of skill before engaging in dangerous operations.'[88]

From here, the law similarly applied this more objective approach to driving offences which required greater regulation following the recent advent of the motor vehicle, and the offence of driving recklessly or negligently became the first explicit attachment of recklessness to a statutory offence.[89] By the introduction of the Road Traffic Act 1930, commentators observed that it was 'hard to see what "reckless" can mean except "realising the possibility of harm to others".'[90] When the courts came to reconsider the definition of recklessness in relation to driving offences some decades later, the Court of

[84] Dori Kimel, 'Inadvertent recklessness in criminal law' (2004) 120(Oct) *Law Quarterly Review* 548, 553.
[85] Sally Cunningham, 'Recklessness: Being reckless and acting recklessly' (2010) 21(3) *King's Law Journal* 445.
[86] *Ibid*., 447.
[87] *R v Williamson* (1807) 172 ER 579.
[88] Cunningham (2010), 447; citing K. J. M. Smith, *Lawyers, Legislators and Theorists: Developments in English Jurisprudence 1800 – 1957* (Clarendon Press 1998), 89.
[89] Motor Car Act 1903, s. 1.
[90] J. W. Cecil Turner, '*Mens rea* and motorists' (1933) 5 *Cambridge Law Journal* 61, 75.

Appeal in *R v Murphy*[91] adopted a notably objective test of recklessness, commenting that 'everybody knows that there is a risk of an accident if a vehicle is not driven with due care and attention on the highway, whether he thinks about it or not.'[92]

Meanwhile, the adoption of recklessness (alongside intention) as an interpretation of "malice" resulted in a contrastingly subjective approach. Malice was generally required for more serious crimes such as those resulting in death (manslaughter) or serious injury, and so the preference for a more subjective approach might reasonably be attributed to the general belief that such serious crimes required proof of a subjective guilty state of mind, whereas less serious crimes such as reckless driving and criminal damage might be proven from an objective test only. Where criminal offences were defined with malice as the requisite *mens rea*, therefore, recklessness 'increasingly assum[ed] a meaning of conscious risk taking' whereby proof was required that the defendant had been subjectively aware of the relevant risk.[93] Recklessness was subsequently defined subjectively by Turner as existing where the defendant has 'foreseen that a particular kind of harm might be done and yet has gone on to take the risk of it,'[94] which would proceed to be adopted in the seminal decision *R v Cunningham*.[95] Viewed from this historical perspective, Cunningham submits that the real difficulties in understanding recklessness began to emerge when the courts attempted to merge two concepts which had evolved from different backgrounds and concerned substantively different offences.[96]

*

As Herring writes, and this thesis affirms, however,

> '[I]t is submitted that neither a purely subjective nor purely objective test for recklessness is adequate. Under the purely subjective approach, those

[91] *R v Murphy* [1980] QB 434.
[92] *Ibid*., 440,
[93] Cunningham (2010), 451; citing Smith (1998), 165; *R v Pembliton* (1874) LR 2 CCR 119; *R v Welch* [1875] 1 QB 23; *R v Martin* [1881] 8 QB 54.
[94] J. W. Cecil Turner, *Kenny's Outlines of Criminal Law* (19th ed. Cambridge University Press 1966), [158].
[95] Cunningham (2010), 452; citing *R v Cunningham* [1957] 2 QB 396.
[96] Cunningham (2010), 452 – 455.

> who fail to see an obvious risk due to their arrogance, drunkenness or indifference to others can escape liability. However, under an objective approach, those who fail to see an obvious risk through no fault of their own (*e.g.,* their age or mental health) can face a conviction.'[97]

In a similar vein, Norrie highlights the mutual deficiencies of entirely subjective or objective approaches to recklessness.[98] On the one hand, 'subjectivism does not go far enough' and, so long as actual awareness of a risk is a prerequisite to responsibility, those people who are callous or entirely indifferent to the potential risks attached to their actions may unreasonably escape liability because of their lack of subjective awareness of a given risk.[99] On the other hand, 'objectivism is too broad' and 'fails to separate the callous from the stupid or merely thoughtless' – it fails to take account of relevant circumstances subjective to the defendant.[100] These challenges to various different conceptions of recklessness are explored in the following discussion of relevant jurisprudence and, as with intention, it shall be demonstrated how each of these cases might have been addressed under the proposed hybrid objective / subjective approach to recklessness.

Once again, inspiration has been taken from the existing legal definition of recklessness in order to arrive at an entirely objective definition. To this extent, the objective definition refers to the existence of *a* (unreasonable) risk but does not specify that such risk must be of a certain likelihood or severity of harm. This follows a consistent approach in jurisprudence which has never placed any particular requirements on the likelihood or severity of a given risk once it is foreseen.[101] Equally, the objective definition only includes *unreasonable* risks within the scope of recklessness, again, as is commensurate with the current law. Herring notes that the requirement that a risk is unreasonable was rarely disputed before the courts because 'it will be unusual for there to be a case where it is reasonable for the defendant to take a risk that a person will be injured.'[102] Nonetheless, such cases do exist; crossing a busy road, overtaking another vehicle,

[97] Jonathan Herring, *Great Debates in Criminal Law* (4th ed. Red Globe Press 2020), 70.
[98] Alan Norrie, *Law and the Beautiful Soul* (GlassHouse Press 2005).
[99] *Ibid*., 83.
[100] *Ibid*., 83 – 84.
[101] *R v Brady* [2006] EWCA Crim 2413.
[102] Herring (2020), 144.

medical surgery and various sports all carry risks of causing harmful consequences, but they are *reasonable* risks to take. Thus, whether or not a risk is reasonable ultimately 'involved a value judgment,'[103] which is made objectively in both the extant law and the proposed hybrid approach to recklessness.

### *10.2.2. Testing Hybrid Recklessness in Jurisprudence*

#### *10.2.2.1. R v Cunningham*

The modern subjective test for recklessness finds its origins in *R v Cunningham*[104] and is frequently referred to as "Cunningham" recklessness in order to distinguish it from "Caldwell" recklessness, considered further below. Neither the facts nor outcome in *Cunningham* are especially complex; the defendant stole a gas meter from the cellar of a house and fractured a gas pipe in the process, failing to turn off a nearby stop tap and causing gas to escape into an adjoining property and bedroom. The victim was sleeping at the time and inhaled a considerable quantity of gas, causing injury but not death. The defendant was convicted *inter alia* of unlawfully and maliciously causing a noxious poison to be administered to another,[105] and the verdict was appealed on the basis that the judge misdirected the jury as to the meaning of "maliciously."

The Court of Appeal allowed the appeal on the basis of the alleged misdirection, and instead provided that malice does not refer to any sense of "wickedness" in general, but requires either '(1) an actual intention to do the particular kind of harm that in fact was done; or (2) recklessness as to whether such harm should occur or not (*i.e.*, the accused has foreseen that the particular kind of harm might be done and yet has gone on to take the risk of it.'[106] In so doing, the Court provided a clearly subjective definition of recklessness dependent upon what a particular defendant actually foresaw. Because the Court could not be sure that a jury would similarly have convicted under the new direction, the relevant conviction was quashed for being unsafe. Notably, the defendant had

[103] Loveless, Allen and Derry (2020), 115 – 116.
[104] *R Cunningham* [1957] 2 QB 396.
[105] Offences Against the Person Act 1861, s. 23.
[106] *Cunningham* [1957], 399 – 400; citing J. W. Cecil Turner, *Kenny's Outlines of Criminal Law* (16th ed. Cambridge University Press 1952), 186; *Pembliton* (1874).

declined to give evidence at trial such that there would have been no direct evidence either way regarding whether or not he had subjectively appreciated the risks arising from stealing the gas meter.

The hybrid approach to recklessness would have no difficulty in upholding the original conviction for, to paraphrase, unlawfully and intentionally/recklessly causing a noxious poison to be administered to another. Beginning with the objective definition of recklessness, clearly stealing a gas meter that is connected to live piping carries *a risk* of causing gas to be released and harming another living in the same building. Moreover, the risk in the present case was patently unreasonable; the defendant was committing an act of theft. What is more, even had this not been the case, for example, because he had been removing his own gas meter, the risk is still unreasonable unless the defendant possessed some particular expertise in gas fittings and maintenance. The release of gas into properties can result in numerous mischiefs, from poisoning inhabitants to causing explosions, and it is for this very reason that people who work with fitting and maintaining gas connections invariably require training and certification. As to the second limb of the test, not least because the defendant declined to offer any evidence, there are no relevant circumstances to suggest that it would not be reasonable to expect any other person in the defendant's place to appreciate that unreasonable risk.

#### 10.2.2.2. *R v Briggs*

In the case of *R v Briggs*,[107] the defendant landlord was moving property out of a garage and driveway on which the victim tenant had two vehicles, including a Mini; there had evidently been previous tension between the parties regarding the tenant's use of the garage. The victim was anxious regarding her vehicles and watched the defendant as he moved various items. At one point, the victim reported seeing the defendant physically push the vehicle so that he could close the garage door and, at another point, she reported seeing the defendant near the vehicle door, making up-and-down motions with his arm, albeit she could not see precisely what he was doing. The victim later discovered that the door handle to her Mini was missing, whilst evidence from the police suggested that it

[107] *R v Briggs* [1977] 1 WLR 605.

would have required considerable force to break the handle. At trial, the defendant first submitted no case to answer, but the trial judge considered that there was a case to be answered and the defendant was called to give evidence. His evidence agreed in large with that of the victim, except that he refuted touching or going anywhere near the car door handle. In the event, the jury explicitly preferred the victim's evidence and convicted the defendant of criminal damage, which may be committed intentionally or recklessly;[108] however, his conviction was overturned on appeal due to misdirection by the trial judge on the meaning of recklessness.

Assuming, as the jury did, that the victim's account of events is to be preferred along with the police evidence, the first question under the hybrid test is whether there was an unreasonable risk that the car door handle would be damaged by the defendant moving property or pulling the handle with force? Certainly, such a risk exists; lifting and moving heavy items in a narrow space by a car could indeed result in damaging the handle, and taking and pulling the handle by force would be yet more likely to cause damage. Moreover, there is no difficulty in concluding that such a risk was unreasonable; if the defendant damaged the handle on purpose there can be little reasonable excuse. And, even if the handle was knocked in the process of moving items, the space was tight and the defendant knew that the victim was home, and he could easily have asked her to move the vehicle first. With regards to the second limb of the hybrid test, the defendant offered no evidence of circumstances that would reasonably excuse anybody from failing to appreciate the risk of damage to the vehicle. Therefore, the hybrid test would uphold the original conviction for recklessly causing criminal damage; and, indeed, from the case report of *Briggs*, it is clear that the jury believed the victim's evidence that the defendant had explicitly approached and done something to the vehicle handle,[109] thus supporting the conclusion of the hybrid test in convicting in this case.

[108] Criminal Damage Act 1971, s. 1.
[109] *Ibid.*, 606.

#### 10.2.2.3. *R v Parker*

Whereas the preceding case is relatively simple, the case of *R v Parker*[110] forced the courts to stretch the subjective test for recklessness almost to breaking point, as the Court of Appeal was clearly very minded that the defendant be convicted on the one hand, but struggled to support such a conviction applying an entirely subjective test on the other. The defendant was witnessed by police officers "smashing down" a telephone onto its receiver at a public telephone kiosk, and found the plastic Bakelite phone to be broken when they went to investigate. The defendant explained how one event after another had gone wrong throughout the day and, when he had failed to properly operate the telephone, he had finally entered into a 'great temper and quite plainly the explanation of the situation was partly his frustration at the series of events which had befallen him that evening and partly in anger at the telephone for failing to operate according to his wishes.'[111] Consequently, whilst the substantive facts were not in dispute, the defendant denied having ever *subjectively* foreseen a risk of damaging the phone by slamming it on the receiver, because he had been in such a temper at that moment in time.

The defendant was convicted of intentionally or recklessly causing criminal damage; however, the trial judge gave a direction on recklessness which appeared to cross into the realm of an objective test, suggesting that a person does something recklessly if they act 'without thought for the consequence of it.'[112] Clearly, "without thought" is not the same as the subjective test from *Cunningham* and *Briggs* which requires that the defendant was *actually* aware of a risk which they proceeded to act against. Referring explicitly to *Briggs*, the Court of Appeal in *Parker* affirmed the subjective test, but then proceeded to highlight the relevant circumstances that the defendant was taken to be fully aware of, including the fact that the telephone was made out of plastic Bakelite and the fact that he brought the phone down onto the receiver with force. The defendant's conviction was thus upheld by the Court of Appeal, with Geoffrey Lane LJ concluding:

[110] *R v Parker* [1977] 1 WLR 600.
[111] *Ibid.*, 602.
[112] *Ibid.*, 603.

> '[I]n those circumstances, it seems to this court that if [the defendant] did not know, as he said he did not, that there was some risk of damage, he was, in effect, deliberately closing his mind to the obvious – the obvious being that damage in these circumstances was inevitable. In the view of this court, that type of action, that type of deliberately closing of the mind, is the equivalent of knowledge and a man certainly cannot escape the consequences of his action in this particular set of circumstances by saying, "I never directed my mind to the obvious consequences because I was in a self-induced state of temper".'[113]

As Herring correctly highlights, a case such as *Parker* demonstrates the problems with an entirely subjective approach to recklessness; the more thoughtless or inconsiderate a person is – the more they "close their mind" to obvious risks – the less likely they are to be proven to have been *subjectively* aware of any particular risk. If the Court of Appeal had not upheld the defendant's conviction, 'the kind of claim Parker was making could be made by many defendants' in order to escape liability.[114] Further, as Herring analogises, it would arguably be rare that a defendant punching their victim on the nose would consider in the same moment "my action might harm the victim." However, the Court of Appeal's reasoning is not entirely convincing within a subjective test; they are asserting, in essence, that the defendant was reckless 'because he *ought to have known* about the risk, whether he did in fact know about the risk or not.'[115] This is an objective approach to recklessness couched within the terms of "closing one's mind to the obvious" or risks being "so obvious they must have been at the back of one's mind."

The hybrid test of recklessness would have no difficulty convicting in the circumstances of *Parker*. Clearly there is a risk of damaging a plastic Bakelite phone by smashing it heavily on the receiver, and there are few circumstances where risking damage to public property in such a way could be considered *reasonable*. Turning to the second limb of the test and taking account of the defendant's assertion that he was in a frustrated temper, the

[113] *Ibid*., 604.
[114] Herring (2020), 57.
[115] *Ibid*.

relevant question is *whether it is reasonable to expect anybody in a temper to appreciate the unreasonable risk of damaging a plastic phone by slamming it on its receiver?* For the reasons given by Herring and, indeed, the Court of Appeal in *Parker*, the answer to this question must be affirmative. It is perfectly reasonable to expect members of society to appreciate the risks of how their actions might damage public or other peoples' private property and, moreover, people are not generally excused of such actions merely because they are frustrated, angry, or are having a bad day. Thus, whereas the purely subjective approach struggles to reach the correct verdict in this case without significantly stretching the meaning of subjective recklessness, the hybrid approach has little difficulty in justly and adequately resolving this case.

#### *10.2.2.4. R v Stephenson*

The challenges revealed in *Parker* notwithstanding, the core advantages of a subjective approach are equally revealed in the case of *R v Stephenson*.[116] The defendant in this case was homeless and suffered from schizophrenia, and one evening went to sleep in a large straw stack in a field. He hollowed out one side of the stack and attempted to sleep in the space but, feeling too cold, he lit a fire of sticks and straw inside the hollow of the straw stack. Unsurprisingly, the stack caught fire and caused some £3,500 in damage. The defendant initially told the police that he had caused the fire accidentally with a cigarette, before admitting the next day that it had been an accident resulting from his attempts to stay warm. The defendant gave no evidence at trial, however, and the only witness called for the defence was an experienced consultant psychiatrist who provided evidence as to the defendant's long history of schizophrenia, which could impact upon his ability to appreciate the otherwise obvious danger of lighting an open fire near a straw stack.

The defendant was initially convicted of burglary – which was admitted by the defendant – and arson,[117] and sentenced with a probation order for three years with the condition of submitting to medical treatment. The Court of Appeal considered a number of earlier

[116] *R v Stephenson* [1979] 1 QB 695.
[117] Criminal Damage Act 1971, ss. 1(1) & (3).

cases which appeared to suggest an objective test for recklessness;[118] however, mindful of the existence of a subjective test for recklessness in the law of tort,[119] the Court considered it would be anomalous if an objective test was adopted in criminal law which would be harsher towards the accused than a defendant in tort.[120] The Court of Appeal, therefore, affirmed the subjective approach from *Cunningham* and overturned the conviction for arson, *inter alia*, on the basis that the trial judge misdirected the jury on recklessness and, moreover, failed to adequately explain the potential relevance of the defendant's schizophrenia. Notably, however, the Court explicitly upheld the defendant's sentence of a three-year probation order with submission to medical treatment. With regards to the potentially conflicting *dicta* of *Parker*, the Court of Appeal in *Stephenson* added:

> '[T]he fact that the risk of some damage would have been obvious to anyone in his right mind in the position of the defendant is not conclusive proof of the defendant's knowledge, but it may well be and in many cases doubtless will be a matter which will drive the jury to the conclusion that the defendant himself must have appreciated the risk. The fact that he may have been in a temper at the time would not normally deprive him of knowledge or foresight of the risk. If he had the necessary knowledge or foresight and his bad temper merely caused him to disregard it or put it to the back of his mind not caring whether the risk materialised, or it if merely deprived him of the self-control necessary to prevent him from taking the risk of which he was aware, then his bad temper will not avail him.'[121]

As Fruchtman writes, the eventual outcome in *Stephenson* was undoubtedly correct, but the reasoning is obtuse.[122] The defendants in both *Parker* and *Stephenson* each claimed that the risks of causing damage had not entered their minds at the time of committing the

[118] *Stephenson* [1979], 701 – 702; citing *Andrews v Director of Public Prosecutions* [1937] AC 576, 583; *R v Bates* [1952] 2 All ER 842, 845; *Shawinigan Ltd. v Vokins & Co. Ltd.* [1961] 1 WLR 1206, 1214.
[119] *Herrington v British Railways Board* [1971] 2 QB 107.
[120] *Stephenson* [1979], 703.
[121] *Ibid*., 703 – 704.
[122] Earl Fruchtman, 'Recklessness and the limits of *mens rea*: Beyond orthodox subjectivism: Part I' (1987a) 29(3) *Criminal Law Quarterly* 315, 331.

offences – the claim in each case was that the defendant was *not* subjectively aware of the risk and, therefore, *not* reckless. Meanwhile, the court was clearly mindful not to let temper or rage interfere with a finding of subjective recklessness in *Parker*, whilst entirely accepting that schizophrenia might so interfere in *Stephenson*. The Court of Appeal attempted to explain this difference in treatment by reference to whether, in the case of *Parker*, certain knowledge had first entered the mind before being driven out by anger, or whether, in the case of *Stephenson*, a condition such as schizophrenia prevented such knowledge from ever entering the mind in the first place. Instead, Fruchtman suggests that 'what the court is really saying is that although neither Parker nor Stephenson actually adverted to the risk of harm caused by their actions, Parker had the ability to do so and Stephenson did not.'[123] Thus, the defendant in *Parker* had the capacity to apply their relevant knowledge to their actions and foresee the harm that might result, but was prevented from doing so due to his emotions; conversely, the defendant in *Stephenson* might have had access to such relevant knowledge, but lacked the capacity to apply that knowledge in the event. This explanation closely describes how the hybrid objective / subjective approach to recklessness would deal with the facts of *Stephenson*.

Starting at the first limb of the test, it may readily be stated that lighting a fire in a hollowed-out straw stack carries an unreasonable risk of causing damage by fire. Both the likelihood and severity of damage are obviously high whilst there can be little excuse in the circumstances as to why lighting such a fire would be objectively reasonable. With regards to the second limb of the test, the relevant question is *whether it is reasonable to expect anybody suffering from the effects of schizophrenia to appreciate the unreasonable risk of causing damage by lighting a fire in a hollowed-out straw stack?* As is suggested by Fruchtman, what this second limb of the test is particularly concerned with is the existence or otherwise factors or circumstances that would be accepted as mitigating or eradicating entirely anybody's capacity to make the relevant appreciation between their actions and consequences – it is concerned, *inter alia*, with the defendant's capacities. The defendant might utilise this second limb in two ways.

[123] *Ibid.*, 331 – 332.

First, the defendant might simply deny hybrid recklessness outright, *i.e.,* denying that, due to some relevant factors such as his schizophrenia, it would not be reasonable to expect anybody in the same circumstances to appreciate the unreasonable risk of their actions. In the instant case of *Stephenson*, it is relevant on the one hand that the defendant did not provide any evidence at trial as to their state of mind and, notwithstanding the rejection of an entirely subjective test, evidence of whether or not the defendant *actually* foresaw harm would nonetheless have evidential significance. It is not being schizophrenic *per se* that can absolve someone of responsibility, but the *effects* of such schizophrenia at the relevant moment when an offence was being committed. Thus, whether or not the defendant was suffering from those effects at the relevant time is of evidential value.

What evidence the defendant did adduce was that of an experienced consultant psychiatrist who attested that the defendant's condition 'would have the effect of making the appellant quite capable of lighting a fire to keep himself warm in dangerous proximity to a straw stack without having taken the danger into account.'[124] To lightly paraphrase the Court of Appeal's closing remarks:

> '[T]he mere fact that a defendant is suffering from some mental abnormality which may affect his ability to foresee consequences or may cloud his appreciation of risk does not necessarily mean that on a particular occasion his [capacity for] foresight or appreciation of risk was in fact absent. In the present case, for example, if the matter had been left to the jury for them to decide in the light of all the evidence, including that of the psychiatrist, whether the appellant [possessed the capacity to] have appreciated the risk, it would have been open to them to decide that issue against him and to have convicted.'[125]

On balance, the prosecution might have struggled to prove beyond doubt the reasonableness of expecting anybody in the defendant's circumstances to appreciate the

[124] *Stephenson* [1979], 699.
[125] *Ibid*., 704.

risks of their actions. Conversely, the defendant may have disadvantaged themself by failing to adduce crucial evidence of whether or not their schizophrenia was a factor *in play at the time of the offence*. Therefore, whether or not (a) the defendant's schizophrenia was a relevant factor *at the time of the offence*, and (b) it is reasonable to expect anybody in a similar position to appreciate the risks of their actions, are the crucial questions for the jury to decide. It is submitted that this could be determined either way according to the values and standards of the jury, which represents a point in legal decision-making where the standards and sensibilities of the individual jurors will determine the limits of legal action, for which very reason trial by jury is so valued. That having been said, it is suggested that the prosecution would indeed struggle to meet their burden of proving hybrid recklessness *beyond reasonable doubt*, having regard to the expert psychiatrist's evidence.

The second approach open to the defendant would be to attack the underlying presumption of his capacity for reasons responsiveness and ordinary self-control. In this regard, the defendant would raise the defence of temporary insanity by reason of his schizophrenia which, for example, Glanville Williams opines would have been 'perfectly clear on principle.'[126] The interaction of legal defences with the hybrid approach to *mens rea* is more fully described in chapter eleven of this thesis, below. For present purposes, however, if the defence of insanity on grounds of the defendant's schizophrenia was credibly made out, this would undermine the presumption that, at the relevant time, the defendant possessed the capacity for reasons responsiveness and / or ordinary self-control. Consequently, it is most likely that the defendant in *Stephenson* would not be found guilty under the hybrid approach to recklessness which, it is submitted, is widely accepted as the correct outcome in the present case. Of course, that is not to say that the defendant could not still be made subject to an order for medical treatment, as would likely be the case under the extant law following a successful insanity plea.

[126] Glanville Williams, 'The unresolved problem of recklessness' (1988) 8(1) *Legal Studies* 74, 78.

#### *10.2.2.5. Commissioner of Police of the Metropolis v Caldwell*

In the infamous case of *Commissioner of Police of the Metropolis v Caldwell*,[127] the House of Lords turned definitively towards an objective definition of recklessness. Complicating the matter, however, was the fact that the new, objective *Caldwell* test was applied for relatively new offences 'wherever a statute used the word "reckless",' but the older, subjective *Cunningham* test continued to be applied to 'older statutory offences defined by "maliciously".'[128] Complicating matters further still, the House of Lords applied the objective test to the offence of manslaughter in *R v Seymour* in 1983,[129] only to later disapprove of this extension in *R v Adomako* in 1995.[130] Similarly, the objective test was applied to other offences such as common law assault and criminal damage, taking even longer for these offences to be similarly reverted back to the subjective test.[131]

The defendant in *Caldwell* was an employee at a residential hotel and bore a grudge against his employer. Drunk one night, the defendant decided to set fire to his place of employment, and was charged with intentionally or recklessly causing criminal damage[132] and damaging property whilst intending to endanger life or being reckless as to whether the life of another would be so endangered.[133] The defendant pleaded guilty on the first charge but contested the latter more serious charge; it was the defendant's evidence that he had been so drunk at the time of the offence that it did not occur to him at all that there might be people inside the hotel whose lives would be endangered by setting the building alight. At first instance, the trial judge directed the jury that self-induced drunkenness provided no defence against the charge of intentionally or recklessly endangering life, and the defendant was subsequently convicted and sentenced to three years' imprisonment. The Court of Appeal quashed the conviction for the contested aggravated charge but maintained the same three-year sentence with respect to the former charge on which the defendant pleaded guilty.[134] Specifically, the Court considered that

---

[127] *Caldwell* [1982] AC 341.
[128] Loveless, Allen and Derry (2020), 119 – 120.
[129] *R v Seymour* [1983] 2 AC 493.
[130] *R v Adomako* [1995] 1 AC 171; see John E. Stannard, 'From Andrews to Seymour and back again' (1996) 47(1) *Northern Ireland Legal Quarterly* 1.
[131] Loveless, Allen and Derry (2020), 119.
[132] Criminal Damage Act 1971, s. 1(1).
[133] *Ibid*., s. 1(2).
[134] *R v Caldwell* (1980) 71 Cr App R 237.

the trial judge had misdirected the jury that drunkenness was not a relevant matter to be considered, because the aggravating element of endangering life was one of specific intent.

Appeal to the House of Lords by the Commissioner of Police was dismissed by a majority of three-to-two, with Lord Diplock providing the judgment for the majority. He considered broadly that there was no difference in blameworthiness between the person who takes a risk because, having crossed their mind, they have dismissed and proceeded to take that risk in rage, excitement or drunkenness, and the person who takes the same risk because it did not enter their mind at all. What is more, Lord Diplock considered that it would not be practicable to ask the jury to make such a distinction in any event; in a statement that encapsulates the core issue at the heart of subjective *mens rea*, he says 'the only person who knows what were the accused's mental processes is the accused himself – and probably not even he can recall them accurately when the rage or excitement under which he acted has passed, or he has sobered up.'[135]

Considering that Parliament had intended the "revise" the law through the Criminal Damage Act 1971 rather than "perpetuate" old confusions from the prior law under the term "malice", Lord Diplock considered that the new statutory term of recklessness must include both recognising and acting regardless of a risk *and* acting without giving any recognition or regard whatsoever for a risk.[136] He thus provided the model direction, that a person is reckless if:

> '(1) he does an act which in fact creates an obvious risk that property will be destroyed or damaged and (2) when he does the act he either has not given any thought to the possibility of there being any such risk or has recognised that there was some risk involved and has nonetheless gone on to do it.'[137]

[135] *Caldwell* [1982], 351 – 352.
[136] *Ibid*., 353 – 354.
[137] *Ibid*., 354.

This was not, however, determinative of the case, which instead turned on the application or otherwise of the defence of intoxication. Lord Diplock affirmed the previous decision of the House of Lords in *Majewski* – that self-induced intoxication may be relevant to crimes of specific intent but provides no defence to crimes of basic intent; however, he considered that the reference to specific and basic intent was unhelpful. Put more simply, self-induced intoxication is 'no defence to a crime in which recklessness is enough to constitute the necessary *mens rea*.'[138] The appeal was dismissed because the Court of Appeal had reached the correct conclusion; notably, both the House of Lords and Court of Appeal nonetheless maintained the defendant's conviction of three years' imprisonment, reflecting a clear condemnation of his actions. Had the trial judge given the correct direction regarding the crucial issue of the defence of intoxication, it is almost certain that the defendant would have been convicted of the aggravated offence (as in the original trial), because his drunkenness would provide no defence to the charge of recklessly endangering life.

It is equally certain that the defendant would be convicted under the hybrid approach to recklessness. Regarding the first limb, there is no question that there is an unreasonable risk of endangering life by setting fire to a residential hotel; the risk is, indeed, arguably great, whilst there can be no objectively reasonable justification for taking such a risk. Concerning the second limb, only the defendant's intoxication was presented as a circumstance relevant to the question, which becomes *is it reasonable to expect anybody intoxicated through alcohol to nonetheless appreciate the risk of endangering life by setting fire to a hotel?* Patently, the answer to this question must be yes – it would not widely be considered reasonable for the law to excuse the taking of all unreasonable risks simply because an individual had become drunk. Society expects a higher standard of conduct, and those liable to risky behaviour when intoxicated are arguably expected to display greater moderation and self-control. That is not to say, however, that *addiction* to alcohol ought not to be treated differently to the defence of mere intoxication, which is a proposal that is explored in greater detail in chapter eleven of this thesis, below.

[138] *Ibid.,* 355.

It is notable to highlight that, had the aggravated charge in *Caldwell* solely been an intention to endanger life, the second limb of the test might be answered differently. It is perhaps more reasonable to expect that somebody who is highly intoxicated might not appreciate a *virtual certainty* of endangering life, as opposed to appreciating simply a risk. This, therefore, reflects the different treatment of self-induced intoxication to offences of specific and basic intent in the extant law. Moreover, the application of the hybrid approach to *Caldwell* has crucially demonstrated how the same objective / subjective approach to recklessness could appropriately deal with the circumstances of both *Cunningham* and *Caldwell* – and, indeed, the cases falling therebetween – without the need for revising the definition of recklessness.

Crosby suggests that a hybrid approach that takes into account a defendant's particular capacities to foresee given risks would not only have provided the appropriate solution to remedy grievances against *Caldwell*, but may have ultimately been intended in the first place.[139] *Caldwell* was not necessarily a case about the definition of recklessness *per se*, but was more intimately concerned with the availability and application, or otherwise, of the defence of intoxication to offences of which recklessness may comprise the *mens rea*. Lord Diplock cites with approval the current approach to voluntary intoxication, namely that an individual who is unaware of a risk due to self-induced intoxication is considered to foresee those risks 'of which he *would have been aware* had he been sober.'[140] As Crosby argues, this suggests that a 'capacity to foresee risk could have been an essential element of the objective test' under *Caldwell* all along.[141] Thus, whereas it was intended that those who failed to foresee risks due to drunkenness, rage or other emotional excitement would fall within the *Caldwell* definition of recklessness because they otherwise possessed the capacity to so foresee such risks, 'it is less clear that [the court] really intended that individuals who were incapable of ever foreseeing the risk could be guilty of an offence.'[142]

[139] Cath Crosby, 'Recklessness – The continuing search for a definition' (2008) 72(4) *Journal of Criminal Law* 313.
[140] *Ibid*., 318.
[141] *Ibid*.
[142] *Ibid*.

Crosby further develops these ideas to suggest her own hybrid approach to recklessness with some notable similarity to the proposal in this thesis.[143] Specifically, she writes:

> 'A broader, capacity-based approach to recklessness, taking into consideration [the defendant's] cognitive capacity and knowledge at the time of the *actus reus*, including *why* [he] failed to see the risk or continued to act despite his appreciation of it, is advocated… As far as the capacity of [the defendant] is concerned, where there is evidence of incapacity exculpation will be dependent upon the extent to which such incapacity is manifest in the circumstances and the extent to which this is fault-free.'[144]

Crosby's description captures many aspects of the second limb of the hybrid approach to recklessness supported in this thesis. In particular, this limb is asking whether or not there exist any circumstances of a given defendant which would be relevant to whether it is reasonable to expect anybody in the same circumstances to appreciate the risks of their conduct. Furthermore, the legal presumption underlying the hybrid approach to *mens rea* in general is the presumption that adults possess the capacities to be responsive to reasons and for ordinary self-control. Together, therefore, these aspects of *mens rea* are fundamentally concerned with capacities – cognitive capacities in particular – that the law has determined to be relevant to the question of legal responsibility. In practice in the courtroom, these capacities are explored and often denied through the application of legal defences such as insanity, automatism or even duress; each of which can be related to one of the aforementioned capacities of realising the consequences of their behaviour, of reasons responsiveness, and of ordinary self-control.

In a similar vein, the seminal jurist Glanville Williams provides a narrower interpretation of *Caldwell* which also reflects aspects of the hybrid approach supported in this thesis.[145] Where Lord Diplock referred to an "obvious risk" in the key passages of his judgment, he earlier expressed that recklessness 'presupposes that if thought were given to the matter

[143] Cath Crosby, 'Gross negligence manslaughter revisited: Time for a change of direction?' (2020) 84(3) *Journal of Criminal Law* 228.
[144] *Ibid*., 235 – 236.
[145] Glanville Williams, 'Recklessness redefined' (1981) 40(2) *Cambridge Law Journal* 252.

by the doer before the act was done, it would have been apparent to him that there was a real risk of its having the relevant harmful consequences.'[146] Williams refers to this as a 'conditionally subjective test', as it enquires what a particular defendant subjectively *would* have thought had they given consideration to their actions.[147] This narrows the interpretation of *Caldwell* from simply asking what risks a reasonable man would foresee, because it accounts for the defendant's relevant circumstances to ask what it is reasonable to expect that *he* would have foreseen. Amirthalingham takes this analysis one step further, suggesting that 'the proper question in any case is not whether the defendant *would* have considered the risk but whether the defendant, having chosen to act in that way, *should* have considered the risk.'[148] Amirthalingham's alteration changes the question from one of hypothetical inquiry into the defendant's subjective state of mind to a question of whether their conduct has breached an objective normative standard set by the criminal law. This captures the essence of the hybrid approach proposed in this thesis.

The criminal law provides proscriptive rules against certain conduct, described by the *actus reus* of any given offence; *mens rea* serves to determine whether a particular defendant is responsible for conduct which breaches that *actus reus*. It achieves this by asking whether that conduct fits within one of a number of types of objectively defined *mens rea*, and whether the law – or, indeed, the jury – considers it reasonable to expect that anybody sharing the defendant's relevant circumstances would appreciate the consequences of their actions, as defined by the particular types of *mens rea* under consideration. Viewed from this perspective, *mens rea* under the hybrid approach is asking neither *what* a particular defendant subjectively had in their mind, nor what they *would* have had in their mind had they given appropriate consideration; but is asking what the criminal law – the jury – considers any reasonable defendant *should* have had in their mind. This, it is again submitted, captures part of what is being expressed by Williams' narrower interpretation of *Caldwell*, as altered by Amirthalingham.

[146] *Caldwell* [1982], 351.
[147] Williams (1981), 269 – 270.
[148] Kumaralingam Amirthalingam, 'Caldwell recklessness is dead, long live *mens rea*'s fecklessness' (2004) 67(3) *Modern Law Review* 491, 494.

#### 10.2.2.6. *R v Lawrence*

Decided just after *Caldwell* but with judgment released the same day, the House of Lords provided a similar substantive decision on the interpretation of recklessness in *R v Lawrence*,[149] with Lord Diplock again providing the leading judgment of the Court. The defendant in this case was driving a motorcycle through a busy urban area at night when he hit and killed a pedestrian, and he was charged causing death by reckless driving.[150] The defendant asserted that he was driving around 30, and no more than 40 miles-per-hour, whilst the prosecution contested that he must have been travelling between 60 and 80 miles-per-hour. The determination of this fact would effectively dispose of the matter, as 'even the defence did not suggest that any sensible jury could come to any other conclusion than that he was driving recklessly' whereas, if the jury determined that the defendant had been driving at the lower speeds, the prosecution had offered no further evidence other than excessive speed in order to establish their case of recklessness.[151] The jury convicted by majority; however, the trial judge's direction was admittedly confusing and verbose, on which basis the defendant's prosecution was ultimately overturned by the Court of Appeal and House of Lords.

That notwithstanding, the majority of the jury had found the prosecution's evidence more credible, concluding that the defendant had indeed been driving at excessive speed at night in an urban area with a 30 mile-per-hour speed limit. Accepting the jury's finding, there is little doubt that the defendant would be convicted under the hybrid approach to recklessness. First, driving above the speed limit in an urban area at night unquestionably carries a risk of causing death and, absent of some specific defence such as necessity, there is no reason to argue that such a risk is reasonable; it is for this reason that speed limits are imposed. Second, there were no relevant circumstances presented by the defence to suggest that it is not reasonable to expect anybody in the same circumstances to appreciate the risk of causing death by speeding through a residential area at night.

[149] *R v Lawrence* [1982] AC 510.
[150] Road Traffic Act 1972, s. 1.
[151] *Lawrence* [1982], 521 – 522.

Clearly, therefore, the defendant's driving was reckless, concurring with the defendant's original conviction in *Lawrence*.

One particular criticism arising against *Caldwell* recklessness that is revealed in particular by *dicta* from *Lawrence* is the so-called Caldwell "loophole." As Birch explains, a loophole exists because the defendant may argue that they had considered and then wrongly, even unreasonably dismissed that risk.[152] This would not fit the entirely subjective definition from *Cunningham*, as the defendant has not simply dismissed a risk but has considered and wrongly concluded its veracity; nor does it fit the objective definition from *Caldwell*, because the defendant has not failed to consider a risk that the reasonable person otherwise would. The loophole is highlighted in *Lawrence* in particular where Lord Diplock permits that, 'if satisfied that an obvious and serious risk was created by the manner of the defendant's driving, the jury are entitled to infer that he was in one or other of the states of mind required to constitute the offence and will probably do so; *but regard must be given to any explanation he gives to his state of mind which may displace the inference*.'[153]

The loophole is an issue because it results in entirely counterintuitive results. Suppose, for example, that Caldwell claimed that he had considered the risk of endangering life but concluded that nobody could possibly be in the hotel because the lights were off. Intuition does not suggest that Caldwell ought to be acquitted, but such a hypothetical case would fall within the loophole. Similarly, Lawrence might claim that he considered the risk of causing injury by his driving, but concluded that his skills were so sufficiently advanced that the risk was negligible. Again, the intuition is that such an excuse should not permit an acquittal yet, in each of these hypotheticals, the defendant has neither become aware of but dismissed a risk out of hand nor failed to become aware of that risk whatsoever.

Moreover, the loophole is not entirely academic; although explicit use of the loophole has not yet been reported, there have been some 'close calls.'[154] In *M. J. J. (a minor) v*

[152] Diane J. Birch, 'The foresight saga: The biggest mistake of all?' (1988) (Jan) *Criminal Law Review* 4.
[153] *Lawrence* [1982], 527.
[154] Birch (1988), 15.

*Cooper*,[155] the Divisional Court quashed the defendant's conviction because, having found a risk to be obvious, the Magistrate's Court failed to proceed to consider whether the defendant was in one of the two definitions of recklessness. As Birch writes,

> '[T]he suggestion that the state of mind of one who genuinely believes his conduct to be safe is excluded is supported by the conclusion reached, which is that [the defendant's] conviction should be quashed because insufficient attention was paid to his explanation that he believed that no damage would result.'[156]

Under the hybrid approach to recklessness, the loophole would be addressed under the second limb of the test. Assuming the defendant's assertions were believed as credible, the question would be whether those assertions, as relevant circumstances, are such that it is reasonable to expect that anybody would not appreciate the unreasonable risk flowing from their actions. This, therefore, becomes a question of what such assertions the jury considers to be sufficiently reasonable in order to undermine the appreciation of the relevant risk.

Suppose the defendant in *Caldwell* considered the risk of endangering life, but concluded that there was no such risk because the hotel lights were off. This reason alone could not be accepted as leading any *reasonable* person to the conclusion that there was no risk of anybody being in the hotel or endangered by setting it alight. Suppose instead that the defendant reached their conclusion because the hotel was closed for business, boarded up and long-abandoned; this, it might be argued by the defence, potentially leads to the conclusion that it is *not* reasonable to expect anybody to appreciate the risk of endangering life by setting fire to an abandoned and boarded-up building. Although, of course, the prosecution might still argue the risk that the building had been occupied by squatters; the matter is left for the jury.

---

[155] *M. J. J. (a minor) v Cooper* (unreported) (CO/1551/84, 2 July 1987).
[156] Birch (1988), 15; further citing *Chief Constable of Avon and Somerset v Shimmen* (1987) 84 Cr App R 7; *R v Bell* [1984] 3 All ER 842.

Suppose, in *Lawrence,* the defendant submitted that they considered the risk of causing death by their driving, but ultimately dismissed that risk because their level of driving skill was so high. Plainly, it would not be reasonable to forgive people failing to appreciate the risks of dangerous driving simply because they believed themselves to be so exceptionally skilled. But suppose instead that the defendant had broken into a racetrack at night to enjoy driving at speed, and had just so happened to cause death by colliding with another who had decided to do the same. Disregarding any other questions of trespass or the objective reasonableness of risk in such circumstances, the defendant might successfully argue that it is not reasonable to expect anybody to appreciate the risk of endangering life by driving around a closed racetrack at night. It appears, therefore, that the Caldwell loophole does not emerge under the hybrid approach to recklessness, because the *reasonable* impact of any relevant circumstances is considered under the second limb of the test.

Fruchtman similarly considers Lord Diplock's *dictum* in *Lawrence* concerning the regard which must be given to any explanation of the defendant's state of mind,[157] and reaches conclusions sharing a number of features of those reached by Williams and Crosby in relation to *Caldwell*, discussed in the previous section of this thesis, above. Reciting the relevant passage from *Lawrence*, Fruchtman writes that whilst it follows that 'proof of *actus reus* will virtually decide the issue of *mens rea* in the case of an ordinary accused, it remains open to the accused to offer evidence to demonstrate that he is not an ordinary person.'[158] He similarly suggests that *Caldwell* recklessness may in fact be narrowed from an entirely objective test owing to the relevant passage from *Lawrence*, again indicating that where the defendant has failed to consider at all a potential risk of their actions, it may be relevant to inquire as to *why* they have so failed before concluding recklessness – *i.e.,* considering the defendant's *capacities* for having realised the relevant risk. This, again, shares similarities with the hybrid approach advocated in this thesis; the second limb of the test in particular is concerned with relevant circumstances of the defendant

[157] Earl Fruchtman, 'Recklessness and the limits of *mens rea*: Beyond orthodox subjectivism: Part II' (1987b) 29(4) *Criminal Law Quarterly* 421.
[158] *Ibid*., 427 – 428.

which would speak to anybody's capacities to appreciate the relevant consequences of their actions, for reasons responsiveness, and for ordinary self-control.

Notwithstanding further criticisms of *Caldwell* recklessness that are explored in relation to subsequent cases, below, the attraction of an objective approach can undoubtedly be found in cases of driving offences such as *Lawrence*.[159] As Cowan writes, there are countless examples on the roads where drivers act with little or no thought towards a potential risk, such as at times when overtaking another vehicle or trying to make a gap in the traffic. This is not because people are entirely unaware of such risks *per se*, but because 'on the roads decisions to act… were often split-second which might be taken virtually without any thought', almost as more of a reflex in response to the surrounding road conditions.[160] If the offence of reckless driving is to govern behaviour on the roads, however, it must be able to address those situations where drivers take unreasonable risks without thinking as, 'indeed, that category of recklessness on the roads might well be as prevalent as that in which the driver actually foresaw the risk and decided to disregard it.'[161]

#### *10.2.2.7. R v Seymour*

The case of *R v Seymour*[162] was significant for confirming the application of the *Caldwell* approach to recklessness to the offence of manslaughter, with facts that present little difficulty. The defendant had a quarrel with the victim with whom he was living and, some time later, they met on a road driving in opposite directions, the defendant driving an 11-ton lorry. There was a minor collision and the victim got out of her vehicle to approach the lorry; at this point, the defendant drove his lorry against the victim's car, purportedly only intending to move it out of the way. However, the lorry apparently hit with such force that the car was moved several feet, one of the tyres became dislodged, and the victim was crushed between two vehicles. The defendant expressed his

[159] See also *R v Reid* [1992] 1 WLR 793.
[160] Veronica Cowan, 'Reckless driving – Recklessness not limited to the subjective test – Extends to cases where D fails to give any thought to the possibility of risk – Direction to jury' (1992) (Nov) *Criminal Law Review* 814, 816 – 817.
[161] *Ibid*., 816.
[162] *R v Seymour* [1983] 2 AC 493.

willingness to plead guilty to causing death by dangerous driving, but the prosecution declined to accept this and charged him with the offence of manslaughter owing to the gravity of the case, of which the defendant was duly convicted. The defendant appealed on the basis that, for the offence of manslaughter, judicial direction had to go beyond that in *Lawrence* to include proof that the defendant subjectively recognised the existence of some risk from their actions. This was dismissed by both the Court of Appeal and the House of Lords, confirming the application of objective recklessness from *Caldwell* and *Lawrence*.

Clearly, on the facts presented, the hybrid approach to recklessness would similarly convict in this case. Concerning the first limb of the test, there is evidently a risk of causing death or serious injury by shoving one vehicle with a huge lorry whilst pedestrians walk nearby. And again, absent of some relevant defence such as necessity, there can be little argument that such a risk is reasonable to take, no less so merely to move one vehicle out of the way. Concerning the second limb of the test, there are no relevant circumstances presented by the defendant to suggest why it would not be entirely reasonable to expect anybody to appreciate the risk of causing death from the actions described. Notably in this case, were the defendant's assertions accepted that he only wished to move the victim's vehicle out of the way and did not even contemplate the possibility of harming them, an entirely subjective *Cunningham* test would have acquitted, arguably contrary to the fair and just outcome. Patently, the defendant was reckless (within an ordinary sense of the word) by using an 11-ton lorry to shove another vehicle out of the way, evidently with some considerable force, and a conviction for manslaughter is entirely appropriate. The objective test reaches this conclusion where an entirely subjective test does not; as demonstrated, however, the hybrid test also reaches the correct conclusion, whilst avoiding the historical to-and-fro between entirely subjective or objective conceptions of *mens rea*.

#### *10.2.2.8. Elliott v C*

The real challenges to the objective *Caldwell* approach to recklessness emerged in *Elliott v C*[163] and the closely following case of *R v R (Stephen Malcolm)*,[164] discussed below. The defendant in *Elliott* was a 14-year-old girl described as being of low intelligence. Having spent the night outside, she entered a wooden garden shed at 5am, found a bottle of white spirit, poured some on the floor and set it alight in order to keep warm. The fire rapidly flared out of control, the defendant left, and the fire continued to destroy the shed and its contents. The defendant was subsequently charged with recklessly causing criminal damage.[165] However, the juvenile court – explicitly mindful of *Caldwell* – considered that regard must be given to the defendant's age and low intelligence, evidenced *inter alia* by her placement in remedial class in school and her difficulty with answering questions in court; to the fact that she was tired, cold and exhausted on the night in question; and to her lack of experience in handling white spirit. Consequently, they concluded that the risk of burning down the shed would not necessarily have been obvious to the defendant, even if she had given thought to the matter.

On appeal to the Divisional Court, the relevant question was whether the "obvious risk" referred to in *Lawrence* was required to be obvious to a particular defendant or obvious to the reasonable man, notwithstanding that foresight of the risk itself was to be assessed objectively following *Caldwell*. Bound by *Caldwell* and *Lawrence*, the Court determined that the obviousness of any given risk was also to be judged objectively, along with foresight of that risk, thus excluding from consideration any relevant factors subjective to the defendant. The main reason for this conclusion appears to be in order to reconcile or align the objective test for recklessness with the similar approach that is taken in cases of voluntary intoxication. Glidewell J opined the position in law regarding intoxication:

> '[E]ven if it resulted in the defendant not thinking at all as to whether there was a risk that property or life would be endangered, [it] nevertheless did not take him out of the state of mind properly described as "reckless",

[163] *Elliott v C* [1983] 1 WLR 939.
[164] *R v R (Stephen Malcolm)* (1984) 79 Cr App R 334.
[165] Criminal Damage Act 1971, s. 1(1).

> [which] is only consistent in my view with their Lordships meaning by the phrase "creates an obvious risk," creates a risk obvious to the reasonably prudent person.'[166]

It is here that objective recklessness arguably took a calamitous turn. As posited by Birch, Crosby, Fruchtman and Williams, above, it is possible to give a narrower interpretation to *Caldwell* recklessness whereby, rather than asking whether a particular risk would be obvious or foreseeable to the reasonable man, the crucial question is whether or not a given risk would have been obvious or foreseeable to a particular defendant, considering their relevant capacities. Indeed, this comes close to the hybrid objective / subjective formulation of recklessness – and *mens rea* in general – advocated in this thesis. However, the Divisional Court in *Elliott* went in the other direction, holding that the obviousness of any risk under the *Lawrence* direction must also be assessed objectively. Consequently, any such relevant subjective circumstances of the defendant could not be taken into consideration when addressing recklessness, resulting in clear injustice and substantial academic criticism of the purely objective approach.

It is suggested that the hybrid approach to recklessness would likely acquit in this case, in concurrence with the general sentiment amongst academic commentary. On the first limb of the test, it is plain that pouring and setting alight white spirit on the floor of a shed carries an unreasonable risk of causing damage to that shed. However, on the second limb of the test, there are a number of relevant circumstances which, it may be readily argued, would suggest that it is unreasonable to expect anybody in those same circumstances to have appreciated the unreasonable risk flowing from their actions. Specifically, at 14-years-old the defendant's age is an obviously relevant factor; similarly, her purported low intelligence could readily affect her ability to make any such appreciation; she had no experience with white spirit; and, the young girl was purportedly tired, cold and exhausted on the occasion. Taken together, it is not necessarily reasonable to expect that anybody in the same circumstances would appreciate the risks of burning white spirit in a shed.

[166] *Elliott* [1983], 946; further citing *Lawrence* [1982], 525; *R v Miller* [1983] 2 WLR 539.

#### *10.2.2.9. R v R (Stephen Malcolm)*

The following year, the argument was put in the case of *R v R (Stephen Malcolm)*[167] that certain subjective characteristics such as a defendant's age, intelligence and maturity ought to be taken into consideration when applying the objective *Caldwell* test – *i.e.,* that whether or not an obvious or serious risk was *foreseeable*, although assessed objectively, ought to be considered from the perspective of what the reasonable person in the defendant's subjective circumstances would have foreseen. The defendant in this case was a 15-year-old boy who committed a series of burglaries and was arrested following a "tip-off" to the police. The boy later went to the house of the person whom he believed had informed against him and threw milk bottles filled with lit petrol against the house windows; he purportedly only had the intention of frightening the informant living in the property and believed that the petrol bombs would burn themselves out in a few minutes. The defendant was charged and convicted, *inter alia,* with arson with intent or recklessness as to endangering life, applying a purely objective definition of recklessness.

The defendant appealed to the Court of Appeal on the grounds that he ought to have been convicted only if his act created a risk that would be 'obvious to someone of his age and with such characteristics as would affect his appreciation of the risk.'[168] The Court dismissed the appeal, however, applying the purely objective *Caldwell* approach to recklessness. The Court considered *inter alia* that the House of Lords refusal of appeal in *Elliott* strongly suggested that that case had been correctly decided, and that therefore:

> '[I]f the risk created by the acts of the accused that the life of another would be endangered is one that must be obvious to any prudent person who had given his mind to it, and the risk of those harmful consequences was not so slight that an ordinary prudent individual would feel justified in treating it as negligible, then, if before doing the act he either fails to give any thought to the possibility of there being any such risk, or having recognised there

[167] *R v R (Stephen Malcolm)* (1984) 79 Cr App R 334.
[168] *Ibid.,* 337.

was such a risk he nevertheless goes on, then recklessness has been established.'[169]

As mentioned above, the cases of *Elliott* and *Stephen Malcolm* serve to demonstrate the deficiencies of an entirely objective test, resulting in considerable academic criticism of *Caldwell* recklessness overall.[170] Field and Lynn write how the purely objective test makes it possible to convict people 'without having had a fair opportunity to make their behaviour correspond with the law,'[171] whilst Ormerod and Laird highlight the particular unfairness of those decisions wherein the defendants' age and learning difficulties 'may have meant that she was incapable of appreciating the risk.'[172] The hybrid approach avoids these issues because, although recklessness is defined and assessed objectively according to what it is reasonable to expect anybody (*i.e.,* the "reasonable man") to foresee, this assessment is done with the defendant's relevant subjective circumstances in mind; the reasonable man becomes the reasonable man *in the defendant's circumstances*.

In the present case of *Stephen Malcolm*, it is first clear that there is an unreasonable risk of endangering life by throwing lit bottles of petrol at somebody's house and windows. With regards to the second limb of the test, and accepting the defendant's submissions that he believed the projectiles would quickly burn out, the question becomes whether it is reasonable to expect anybody of the defendant's age, maturity and understanding to appreciate the risk of endangering life by throwing lit projectiles at a residential house and windows. It is near impossible to give a definitive answer to this question without the jury's benefit of having seen and heard the full evidence from the defence and defendant. Considering the relevant circumstances, however, it is not difficult to argue that there exists *reasonable doubt* as to whether it is reasonable to expect that anybody sharing the defendant's age and maturity would have appreciated the risk of endangering life.

[169] *Ibid*., 338; citing *Miller* [1983].
[170] Mary Seneviratne, 'Carry on Caldwell' (2003) 12(1) *Nottingham Law Journal* 36, 38 – 40; see also John C. Smith, 'Case commentary: *R v Caldwell'* (1981) *Criminal Law Review* 392.
[171] Stewart Field and Mervyn Lynn, 'Capacity, recklessness and the House of Lords' (1993) (Feb) *Criminal Law Review* 127, 128 – 129.
[172] Ormerod and Laird (2020), 113.

#### *10.2.2.10. R v Adomako*

In the case of *R v Adomako*,[173] the House of Lords reversed its earlier position from *Seymour* and declined to apply objective *Caldwell* recklessness to the offence of manslaughter, returning to a subjective *Cunningham* approach. The case also provides an especially interesting application of the hybrid approach to recklessness, as it concerned manslaughter committed by an experience anaesthetist. As shall be demonstrated below, this example reveals how the particular circumstances of a given defendant can both lower *and* raise the standard of reasonableness that is demanded under the second limb of the hybrid test. Beginning with the facts of the case, the defendant was the anaesthetist during an eye operation for which the patient's breathing was maintained by a ventilator. At 11:05 the endotracheal tube of the ventilator become dislodged cutting off the patient's air supply, but the defendant failed to notice this as an alarm on the ventilator had not been turned on. The defendant was alerted to the issue around 11:10 when another alarm indicated that the patient's blood pressure was dropping, and at 11:14 the patient suffered from a cardiac arrest, nine minutes after the tube had initially been disconnected and without the defendant ever identifying the source of the problem.

The defendant was convicted of involuntary manslaughter following what was described in evidence as his 'abysmal' and 'gross dereliction of care', and both his appeals to the Court of Appeal and House of Lords were dismissed. Departing from the previously discussed case of *Seymour*, the House of Lords notably determined that a direction on recklessness in accordance with *Caldwell* and *Lawrence* was no longer required for cases of involuntary manslaughter, seemingly reverting to a subjective *Cunningham* approach.[174] However, the remainder of the case was decided upon principles of gross negligence which is considered in section 10.5.2.6 of this chapter, below. What is interesting in this case is the hypothetical application of the hybrid approach to recklessness to a defendant with skills and experience *higher* than normal, in contrast with previous examples where more focus has been given to a defendant's deficiencies.

[173] *R v Adomako* [1995] 1 AC 171.
[174] Loveless, Allen and Derry (2020), 119.

The first limb of the test is difficult to apply because the case is more properly assessed under negligence, especially as the defendant's criminal action was really an omission to meet the requisite standard of medical care. That notwithstanding, it may be argued that there was an evidently clear risk of causing injury or death to the patient by his anaesthetist improperly attending equipment (*i.e.,* failing to check alarms) and improperly responding to the medical situation (*i.e.,* failing to check the ventilator connections). With regards to the second limb of the test, the fact that the defendant was a qualified and presumably experienced registrar anaesthetist is a relevant circumstance that may have obvious bearing upon whether or not it is reasonable to expect any similarly qualified defendant to appreciate the consequences of their actions. Doctors not only possess a certain expertise in their field but they offer that expertise to patients and undertake responsibility for their medical care. In this instance, therefore, it is perfectly reasonable to have a *higher* expectation that doctors will appreciate the medical consequences of their actions, more so than would be expected of any other non-professional.

In his evidence, the defendant submitted that he had begun to panic somewhat when things in the operation started to go wrong, and he might submit this as a relevant circumstance in an attempt to undermine the aforementioned reasonable expectation. However, even if this explanation is accepted, it is readily arguable that the defendant's special skill, experience and expertise again render it reasonable to expect anybody in the same circumstances to overcome that panic and be mindful of their medical training. Expressed fully, the question under the second limb of the test become, *is it reasonable to expect any registrar anaesthetist who has begun to panic to appreciate the risk of causing injury or death to their patient by failing to properly attend to the patient's medical care?*

Given the general expectation that medical professionals – but anaesthetists and surgeons in particular – can cope with the emergency situations in which their profession places them, it is entirely reasonable to expect that they maintain some level of composure in the face of such situations. The defendant's conduct in *Adomako* not only failed to maintain a minimal level of composure, but was deemed to be abysmal and a gross dereliction of care. Certainly, a higher standard than this is required of people who hold themselves out

to be medical professionals and, in the event, it was entirely reasonable to expect the defendant anaesthetist to appreciate the risks to their patient flowing from their failure to provide adequate care. Put simply, it was reasonable to expect a qualified and experienced anaesthetist to check the relevant alarms on medical equipment in the first place; further, upon becoming aware of the patient's deterioration, it was similarly reasonable to expect a qualified and experienced anaesthetist to check the relevant tubing connections on that medical equipment. Finally, it is reasonable to expect that each of these (relatively basic) actions would be carried out, notwithstanding that the qualified and experienced anaesthetist had begun to panic. The hybrid approach to recklessness would therefore find the defendant responsible, as indeed was the ultimate outcome of the case.

#### *10.2.2.11. R v G*

The final case of recklessness under consideration is *R v G,*[175] which marked the moment when UK law abandoned objective *Caldwell* recklessness and returned to endorsing the subjective *Cunningham* approach. The facts of the case are such that, once again, they presented an unquestionable challenge to the objective approach to recklessness. The two defendants were aged 11- and 12-years and went out camping without their parents' permission. During the evening, the boys entered the back yard of a shop where they set alight some newspapers on the floor; some of the burning papers were thrown under a plastic bin, and the boys left without putting out the fire. The plastic bin caught alight and the fire spread to the shop and rooftops of other nearby buildings, ultimately causing around £1 million in damage. The defendants asserted that they believed that the newspaper would simply burn itself out on the concrete floor of the shop yard, and never thought there was a risk that the fire could spread. They were subsequently charged and convicted of arson applying the *Caldwell* definition of recklessness.

Providing the leading judgment of the House of Lords, Lord Bingham offered four reasons for departing from *Caldwell* altogether and returning to a subjective test for recklessness. First, he relied upon the 'salutary principle that conviction of serious crime should depend on proof not simply that [the defendant] caused (by act or omission) an

[175] *R v G* [2003] UKHL 50.

injurious result to another but that his state of mind when so acting was culpable.'[176] The rejection of this argument has been extensively explored in chapter eight, above, applying the evidence from Part One of the thesis, and is therefore not rehearsed again here. Second, Lord Bingham highlighted the injustice that could be produced by the *Caldwell* approach, not least in the instant case where 11- and 12-year-old children were assessed according to the standards of the reasonable (adult) man. He noted how both the trial judge and jury expressed regret at the original conviction, arguing further that the 'sense of fairness of 12 representative citizens siting as a jury… is the bedrock on which the administration of criminal justice in this country is built.' [177] Undeniably this criticism is true; as demonstrated in this section, an entirely objective approach to recklessness can indeed result in abject unfairness and injustice, albeit this does not necessitate the return to a fully subjective test either.

Third, Lord Bingham was mindful of the considerable academic criticism surrounding *Caldwell* and, although this would not alone be good reason to overturn the law, nor would it be sensible to ignore completely such reasoned and outspoken criticism. Fourth, the judge asserted that *Caldwell* was based on a misinterpretation of the Criminal Damage Act 1971 which, being offensive against principle and resulting in injustice, required redress. The judge further considered whether or not the objective *Caldwell* approach could be adapted in cases involving children such that comparison was made with 'normal reasonable children of the same age'[178] akin to the proposals in this thesis, but he rejected this suggestion for four reasons also. The first and fourth objections are essentially repetitions of the first and fourth arguments against *Caldwell*, above, namely that this would still offend the principle of convicting only upon proof of the guilty mind, and would be a further misinterpretation of the 1971 Act. The second and third objections are related and warrant closer inspection.

Lord Bingham submitted that permitting an amendment of *Caldwell* in relation to children on grounds of their immaturity would be anomalous unless similar amendments were also

[176] *Ibid.,* [32].
[177] *Ibid.,* [33].
[178] *Ibid.,* [37].

permitted for other categories of defendant, such as 'the mentally handicapped on grounds of their limited understanding.'[179] Against this objection, it is simply stated that the present thesis would indeed permit relevant factors other than age to be considered, in which case Lord Bingham's second critique presents no particular issue but, rather, represents the direction in which the law ought to take.[180] However, his third objection follows that 'any modification along these lines would open the door to difficult and contentious argument concerning the qualities and characteristics to be taken into account for the purposes of comparison.'[181] This objection is rejected for a number of reasons. To begin, it is reiterated that the law *already* varies the standard of the reasonable man when it is applied to people with special skills or expertise; if such people may be held to a higher standard of conduct, why cannot people with reduced capacities enjoy the benefit of a lower standard of conduct. Equally, it is reiterated that the law also *already* varies the standard of the reasonable man when it is applied to people who are voluntarily intoxicated; somebody who is drunk is taken to foresee any such risks that they would have foreseen had they been sober.

Indeed, more generally, it is submitted that the vast majority of those such relevant circumstances that would form the basis of comparison under a hybrid approach to *mens rea* are already contemplated within existing, well-defined legal defences. The point is extrapolated more fully in chapter eleven of this thesis, below; but most circumstances that would be relevant to the second limb of the hybrid test are already encapsulated by legal defences, such that there would not be a deluge of 'difficult and contentious argument' concerning which such circumstances are relevant to consider. Furthermore, the relevance of any such circumstances submitted for consideration under the hybrid test is limited by their relationship to the three capacities required for responsibility. The first is the capacity for the defendant to appreciate the consequences of their actions as described by the objective formulation of *mens rea* (*i.e.,* intention as virtual certainty; recklessness as an unreasonable risk; *etc.*), which forms the key question under the second limb of the hybrid test. Second and third are the capacities to be responsive to reason and for ordinary self-control, which form the basis of the presumption of volition underlying

[179] *Ibid*.
[180] See further Crosby (2008), 331 – 332; Kimel (2004), 552.
[181] *R v G* [2003], [37].

*mens rea* in general. Finally, it is submitted that determining which circumstances and factors of a given case are relevant to whether or not it is reasonable to expect somebody to appreciate the consequences of their actions, is an arguably less difficult and contentious challenge than proving beyond reasonable doubt the actual thoughts that a person subjectively held in their mind at the time of committing an offence.[182]

Returning to the instant case of *R v G*, the defendant's conviction was overturned on appeal to the House of Lords, re-affirming the sole use of the subjective test for recklessness from *Cunningham*, and it is submitted that the hybrid approach would similarly acquit in this case. With regards to the first limb of the test, clearly there is a risk of causing damage by fire from leaving burning paper under a plastic bin; what is more, within their context as trespassers and simply burning the paper for fun, the risk was objectively not a reasonable one to take. Turning to the second limb of the test, the defendants' young ages are clearly relevant factors for consideration – indeed, the defendants were only just above the age of ten years at which there is an irrebuttable presumption against criminal legal responsibility due to youth and lack of maturity. It is not, therefore, necessarily reasonable to expect that any young boys of a similar age would appreciate the risk of setting fire to the shop by leaving burning paper outside under a plastic bin; it is readily argued that the defendants in this case could have lacked the capacity to make the critical appreciation regarding the consequences of their behaviour, owing to their youth, immaturity, and consequent lack of understanding about how fire might spread.

### *10.2.3. Final Comments on Recklessness*

Arguably more so than in relation to intention, the courts have particularly struggled with arriving at a single satisfactory approach to recklessness. Jurisprudence has wrestled with competing subjective and objective conceptions of recklessness, even applying different tests to different crimes. On the one hand, a purely subjective test appears to be insufficiently inclusive, excluding those defendants who are so thoughtless, callous or

[182] See similarly Mitchell Davies, 'Lawmakers, law lords and legal fault: Two tales from the Thames River bank: Sexual Offences Act 2003; R v G and another' (2004) 68(2) *Journal of Criminal Law* 130, 144.

simply uncaring that they do not stop to consider risks that would be plain to anybody else. On the other hand, a purely objective test appears to be overinclusive, unfairly punishing people who not only failed to consider a particular risk of their conduct, but who lacked the very capacity needed to recognise and consider that risk in the first place. The challenges may be better illuminated by considering potential defendants for crimes of recklessness as falling within one of three broad categories.

In category A are defendants who foresaw a risk and acted anyway; these defendants pose no difficulty, as they may be convicted on either the subjective or objective test. Further, the law is not principally interested with *why* they proceeded to run a given risk once that risk is foreseen; it does not matter that they considered the risk and then dismissed it as negligible, or thought that they had taken sufficient precautions against the risk. The key excusatory factor in such cases will be whether or not the risk was a reasonable one to take. In category B are defendants who did not foresee a risk, but were perfectly capable of doing so; perhaps they were overly excited or angry in the event, or maybe they simply did not care enough to give thought to the risks of their actions. But crucially, there is nothing to suggest that they would not have been able to foresee any risk had they given such thought. The subjective test of recklessness fails to capture defendants in category B because they have not subjectively contemplated the risk of their behaviour; conversely, an objective test expands the interpretation of recklessness to ensure that defendants in category B cannot escape liability, for example, by pleading that they were in a fit of rage, or that they simply did not care enough to attend to an otherwise obvious risk.

The problem with the objective test is that it also captures defendants in category C who also failed to consider the risks of their behaviour, but because of some relevant reason such as their age, lack of intelligence and immaturity, or the effects of some illnesses such as schizophrenia. A purely objective test that assesses the foreseeability of risk according to the standard of the ordinary reasonable man fails to capture a range of relevant factors and circumstances which undermine the defendant's very capacity to be able to foresee those risks. Therefore, even had the defendant in category C stopped to consider their behaviour, they nonetheless lacked the capacity to necessarily foresee certain risks in the way that an ordinary (or neurotypical) defendant in category B can. What this section of

the thesis has demonstrated is that the hybrid objective / subjective approach to recklessness can satisfactorily include defendants from categories A and B whilst justifiably excluding defendants from category C, which is something that neither a purely subjective nor objective test has been able to achieve.

There has been notable academic support for the modification of objective recklessness into a hybrid test which, whilst continuing to assess recklessness according to what the reasonable man would foresee, takes account of a defendant's *capacities* to foresee a given risk when formulating who the reasonable man actually is. In this way, the law asks what it is reasonable to expect any given defendant to foresee, having regard to their particular characteristics that impact upon the capacity for anybody to foresee risk. To begin, there are countless ways in which people can have different capacities to foresee risk.[183] For example, schizophrenia can seriously interfere with a person's attentiveness and perceptions of what is real, whilst treatment with sedative drugs can further reduce attentiveness (to risk). People experiencing the manic phase of bipolar disorder are often perceived as acting highly recklessly, whilst those experiencing depression often find it more difficult to attend to external events. Meanwhile, evidence has also revealed key differences in the way the even neurotypical people attend to risk when driving, for example, with experienced drivers exhibiting a significantly different pattern of risk-attentive behaviours. Field and Lynn write:

> '[A]ll this suggests that the young, the inexperienced and the mentally disordered may all in different ways lack the capacity to foresee at least some of the risks that the prudent person might perceive as "obvious", or indeed "obvious and serious".'[184]

Concurrently, the law appears to have little difficulty in modifying the reasonable man test when taking into account an individual's special skills, experience or qualifications. For example, where the gross negligence of an anaesthetist is in question, they are judged

[183] Stewart Field and Mervyn Lynn, 'The capacity for recklessness' (1992) 12(1) *Legal Studies* 74, 75 – 76.
[184] *Ibid*., 76.

'by the standards of reasonably competent anaesthetists.'[185] As Jefferson rightly asks, therefore, 'if the standards can be raised in relation to an expert, why cannot it be reduced in respect of an accused who is not ordinary such as the tired, hungry, mentally deficient girl in *Elliott v C*?'[186]

This argument is particularly forceful when considering the relationship between *mens rea* and capacities in general. As has previously been argued throughout this thesis, underlying *mens rea* is the presumption that all adults have the capacity to be responsive to reasons and to exert a degree of ordinary self-control over their behaviour. To this extent, the seminal jurist Hart argues that these such capacities which underlie volitional action are a minimal requirement for holding somebody responsible for their actions and, equally, it is because these capacities are undermined that excusing factors such as involuntariness and insanity may provide defences in law.[187] There is a strong argument, therefore, that certain relevant factors or circumstances that impact upon anybody's capacities to appreciate the consequences of their actions, for reasons responsiveness, and for ordinary self-control, ought to be taken into account when considering *mens rea*, including recklessness. The hybrid approach to recklessness advocated in this section adds this crucial aspect to an otherwise objective test.[188] Having first defined recklessness objectively as acting where there is an unreasonable risk of causing harm, injury or some other adverse (and criminally prohibited) outcome, the test secondly invites evidence of any such relevant circumstances that impact upon the aforementioned capacities, and asks whether or not it is reasonable to expect *anybody* (objective) sharing such circumstances (subjective) to appreciate those risks of their actions.

[185] Michael Jefferson, 'Recklessness: The objectivity of the Caldwell test' (1999) 63(1) *Journal of Criminal Law* 57, 61.
[186] *Ibid*., 62 – 63; see also Barry Mitchell, 'Recklessness could still be a state of mind' (1988) 52(3) *Journal of Criminal Law* 300.
[187] H. L. A. Hart, 'Legal responsibility and excuses' in Hart H. L. A. (ed.), *Punishment and Responsibility: Essays in the Philosophy of Law* (2nd ed. Oxford University Press 2008); see also George P. Fletcher, *Rethinking Criminal Law* (Oxford University Press 2000), 798 – 817; Herbert Packer, *The Limits of the Criminal Sanction* (Stanford University Press 1968), 108 – 113.
[188] See similarly David Ibbetson, 'Recklessness restored' (2004) 63(1) *Cambridge Law Journal* 13.

## *10.3. Knowledge, Belief and Suspicion*

Under the revised hybrid formulation, knowledge is defined objectively as being the *certainty that a particular circumstance exists*, and is assessed with the defendant's relevant subjective circumstances in mind by asking, *is it reasonable to expect anybody in the defendant's circumstances to appreciate the truth that a particular circumstance exists*?

Belief may be understood as 'something short of knowledge.'[189] Under the revised hybrid formulation, belief is defined objectively as being the *conviction that a particular circumstance exists*, and is assessed with the defendant's relevant subjective circumstances in mind by asking, *is it reasonable to expect anybody in the defendant's circumstances to appreciate the conviction that a particular circumstance exists?*

Suspicion falls yet further down the scale below belief.[190] Under the revised hybrid formulation, suspicion is defined objectively as being the *conjecture that a particular circumstance exists*, and is assessed with the defendant's relevant subjective circumstances in mind by asking, *is it reasonable to expect anybody in the defendant's circumstances to appreciate the conjecture that a particular circumstance exists?*

By way of brief explanation, the description of certainty as to the existence of particular circumstances might be regarded as another way of stating, simply, that a particular circumstance is true. In this respect, the objective *legal* definition of knowledge is concerned with whether a given statement as to the existence of circumstances is factually true or not,[191] without delving into the deeper philosophical questions of how knowledge is defined in the broader epistemological sense. To paraphrase a popular judicial pronouncement, "knowledge", "belief" and "certainty" are read with their ordinary and casual meaning.[192] Similarly, where the description refers to the existence of a particular

[189] *R v Hall* (1985) 81 Cr App R 260, 264.
[190] Ormerod and Laird (2020), 131.
[191] *R v Montila* [2004] UKHL 50, [27].
[192] See *R v Saik* [2006] UKHL 18, [26].

"circumstance", this latter word is interpreted broadly to include the existence of facts and things specifically, or of a state of affairs or circumstances more generally.

The definition of belief as a "conviction" denotes that it is something less than certainty, yet still something more than a mere suspicion or "conjecture". Unlike certainty, a conviction may in fact be wrongly held; but it is held with some veracity nonetheless. Beliefs are typically reinforced by some evidence or reasoning which supports the believer in relying on those beliefs. For this reason, beliefs are open to question and may be changed, but it is not characteristically easy to do so. The Court of Appeal authority of *R v Hall* provided that:

> 'Belief, of course, is something short of knowledge. It may be said to be the state of mind of a person who says to himself: "I cannot say I know for certain that [the circumstance exists] but there can be no other reasonable conclusion in light of all the circumstances, in the light of all that I have heard and seen.'[193]

Suspicion is something less still than belief. Suspicion may, again, be incorrectly held, although it is generally held with less veracity than a belief and, equally, is generally supported by weaker evidence or reasoning. But suspicion remains something more than a 'vague feeling of unease', 'inkling' or 'fleeting thought.'[194] The House of Lords in *Hussien v Chang Fook Kam*[195] provided that:

> 'Suspicion in its ordinary meaning is a state of conjecture or surmise where proof is lacking: "I suspect but I cannot prove". Suspicion arises at or near the starting point of an investigation of which the obtaining of *prima facie* proof is the end.'[196]

[193] *Hall* (1985), 264.
[194] *R v Da Silva* [2006] EWCA Crim 1654, [15] & [19].
[195] *Hussien v Chang Fook Kam* [1970] AC 942.
[196] *Ibid*., 948.

### *10.3.1. The Varying Degrees of Knowledge, Belief, and Suspicion*

A number of attempts have been made to categorise different types of knowledge, both in jurisprudence and in academic literature, and in ways that are readily applicable to belief and suspicion also. Moreover, whereas knowledge, belief and suspicion may at first appear to be inherently subjective concepts, as with oblique intention varying degrees of knowledge have been extrapolated in certain circumstances in order to impute knowledge that a particular defendant may not necessarily have subjectively held in the event. What is more, a great many of the offences that can be committed with the *mens rea* of subjective knowledge are tempered by a second alternative *mens rea* element, often posed objectively. For example, it is an offence for one person to sell or transfer any firearm or ammunition to another 'whom he knows or has *reasonable cause for believing* to be drunk or of unsound mind.'[197] Similarly, criminal harassment may be committed when one person pursues a course of harassing conduct that he 'knows or *ought to know* amounts to harassment of the other',[198] whilst numerous further examples of subjective knowledge tempered by a second partially or wholly objective *mens rea* element are forthcoming.[199]

As with other purely subjective forms of *mens rea* and intention in particular, it is relatively rare that a guilty defendant will have plainly and explicitly laid out what they knew, believed or suspected at the time of committing a crime. Rather, more often the jury will have to infer these elements from the defendant's actions, behaviour, and the credibility of their account.[200] All this is to suggest that, again, the proposed move from entirely subjective tests of knowledge, belief and suspicion to a hybrid test that applies

[197] Firearms Act 1968, s. 25.

[198] Protection from Harassment Act 1997, s. 1(1); see also s. 4.

[199] For example, see Psychoactive Substances Act 2016, ss. 5 & 8 – offence of supplying, importing and exporting a psychoactive substance require *inter alia* that the defendant 'knows or suspects, or *ought to know or suspect*, that the substance is a psychoactive substance'; Misuse of Drugs Act 1971, s. 28 – applying to a number of offences, including simple possession of illegal drugs, providing a defence if the accused can demonstrate that they 'neither knew of nor suspected *nor had reason to suspect* the existence of some fact alleged by the prosecution…'; Proceeds of Crime Act 2002, ss. 327 – 329 – offences regarding use of criminal property include *mens rea* elements of knowledge or belief '*on reasonable grounds*'; Terrorism Act 2000, ss. 15 – 18 – a number of offences relating to terrorist funding and property consist of the *mens rea* of subjective intention, knowledge or belief, tempered with a second more objective element of a *'reasonable cause to suspect*' that funds or property will be used for terrorism; Unsolicited Goods and Services Act 1971, s. 4(1) – offence of sending unsolicited publications (containing sexual content) requires *inter alia* that the defendant '*knows or ought reasonably to know*' the material is unsolicited.

[200] *Lee v Taylor and Gill* (1912) 77 JP 66, 69.

objectively whilst incorporating subjective elements, should not be especially controversial. The courts (and juries) are already often engaged with a combination of subjective and objective components relating to the *mens rea* of knowledge, belief and suspicion. This is perhaps most clearly demonstrated through exploring the various "degrees" or types of knowledge that have been considered in jurisprudence and academia.

In the case of *Taylor's Central Garages (Exeter) v Roper*,[201] Devlin J attempts to describe three "degrees" of knowledge, the first two of which are submitted to suffice for the *mens rea* of knowledge.[202] The first degree is actual or direct knowledge, *i.e.,* that knowledge which is at the forefront of a person's mind. However, as Devlin J indicated in *Roper*, 'invariably… it is impossible to prove the state of another man's mind with the result that the defendant's knowledge is generally inferred from the nature of the act done.'[203] Knowledge in the second degree consists of "wilful blindness", a legal concept that has received particular attention within the *mens rea* of knowledge and belief and is explored further in the examples from jurisprudence, below. In brief, however, wilful blindness is described as 'deliberately refraining from making inquiries, the result of which a person does not care to have',[204] and is generally regarded as consisting of a suspicion that a particular circumstance or fact likely exists coupled with a deliberate refusal to investigate the question further through readily available means.[205] The third degree of knowledge is "constructive knowledge", consisting of those things that a person "ought to have known" or, phrased differently, those things that are considered to be known by the "reasonable man". Devlin J asserted that constructive knowledge 'has no place in the criminal law';[206] however, this is clearly no longer the case since the introduction of various statutory

[201] *Taylor's Central Garages (Exeter) v Roper* [1951] 2 TLR 284.
[202] See also John Llewelyn Jones Edwards, 'The criminal degrees of knowledge' (1954) 17(4) *Modern Law Review* 294; Mohamed Elewa Badar, *The Concept of Mens Rea in International Criminal Law: The Case for a Unified Approach* (Hart Publishing 2013), 60 – 61.
[203] *Ibid*., 295; citing *Roper* [1951], 288.
[204] *Roper* [1951], 288.
[205] David Ormerod and Karl Laird, *Smith, Hogan, and Ormerod's Criminal Law* (15th ed. Oxford University Press 2018), 116; Martin Wasik and Mark Thompson, '"Turning a blind eye" as constituting *mens rea*' (1981) 32 *Northern Ireland Law Quarterly* 328, 337 – 341.
[206] *Roper* [1951], 288 – 289.

crimes for which the *mens rea* consists of something that the defendant knew, *ought to have known*, or had *reasonable cause* to believe or suspect.[207]

R. A. Duff similarly separates knowledge into three categories.[208] In the context of discussing *Caldwell* and *Lawrence* recklessness and the knowledge of risks of which a reckless driver might be aware, Duff extrapolates three types of knowledge. First, "explicit" knowledge is that which is '"prominent" in [someone's] mind, to which he consciously adverts',[209] and may be regarded broadly akin to Devlin's first degree of actual or direct knowledge. Second, Duff describes "tacit" knowledge as that which is 'stored in the brain and available if called on' – distinguishing from explicit knowledge which has in fact been called to mind – but which is still readily available for recall and which 'may guide [a person's] actions and reactions without any such conscious process of contemplating his surroundings or calling his latent knowledge to mind.'[210] Third, latent knowledge refers to 'general knowledge… [such as] of the risk which driving involves which would enable [someone] to notice the risks created by his present driving.'[211] Latent knowledge is neither explicitly nor directly in the mind of somebody as they commit a crime, nor is it within the readily available bank of tacit knowledge that influences and guides one's behaviour.

Sullivan provides a clear and illustrative analogy which broadly, albeit not exactly, follows Duff's three categories:

> 'Consider, for example, D, who is smuggling drugs into the United Kingdom for the first time. As he enters the green channel, he is acutely aware of the heroin hidden in his suitcase. We have here the clearest possible case of knowingly evading a restriction on importation.'[212]

---

[207] Badar (2013), 60.
[208] R. Anthony Duff, '*Caldwell* and *Lawrence*: The retreat from subjectivism' (1983) 3(1) *Oxford Journal of Legal Studies* 77.
[209] *Ibid*., 80.
[210] *Ibid*.
[211] *Ibid*.
[212] G. R. Sullivan, 'Knowledge, belief, and culpability' in Shute S. and Simester A. (eds.), *Criminal Law Theory: Doctrines of the General Part* (Oxford University Press 2002), 210.

This is also a clear case of explicit, actual or direct knowledge of the substance that the defendant is carrying and the fact that it is prohibited.

> 'Six months later, D has made many successfully drug-smuggling trips and is now quite relaxed when passing through customs. He gives no thought to the drugs in the case; his mind is on other things. But, again, we are seemingly confronted with a clear case of knowing importation. Were he to be approached by a customs officer, his mind would immediately engage with the drugs in his possession. He has *tacit* knowledge of the presence of drugs and that seems to be enough.
>
> 'Imagine, however, that D is passing through customs some three years later. He has ceased smuggling drugs some two years previously. However, a packet of heroin is to be found in the concealed panel of his case. He had omitted to remove it some considerable time ago and has now forgotten about its presence. The most that can be claimed is that he has *latent* knowledge of the drugs. It is by no means clear, in terms of authority or principle, whether this latent knowledge should suffice as the culpability for the offence of knowingly evading a restriction on importation.'[213]

Whilst the law is relatively settled that both explicit knowledge and tacit knowledge (broadly analogised to wilful blindness) constitute the *mens rea* of knowledge, it is less settled whether latent or constructive knowledge will suffice. What is more, the dividing line between these two may, at times, appear so artificial as to be practically meaningless. For example, the defendant in *R v Bello*[214] was prosecuted for knowingly remaining within the UK after the time permitted by the conditions of his entry, to which he claimed that he had been under significant pressure from various events associated with the recent death of his mother, such that he had 'at no time adverted to the fact that his period of leave had expired.'[215] The Court of Appeal considered that, even accepting the

[213] *Ibid*.
[214] *R v Bello* (1978) 67 Cr App R 288.
[215] Sullivan (2002), 210.

defendant's account of events, no defence was available to him because he would have been capable of recalling the fact that his leave had expired had he considered or been questioned on the matter.[216] In *R v Russell*,[217] by contrast, the defendant was found not to be knowingly in possession of an offensive weapon when, at the time of his arrest in his car, he had entirely forgotten that he had hidden the weapon within the vehicle some months earlier.[218]

The details of these cases are considered further in section 10.3.2, below. For the purposes of the present section, however, the argument is made that the current subjective approach to knowledge, belief and suspicion is heavily moderated by numerous more objective considerations. Whether the *mens rea* of knowledge is combined within particular offences with other more objective *mens rea* elements such as reasonable grounds / cause for belief or suspicion, or what ought to have been known, or whether knowledge is imputed by means of the concepts of wilful blindness or constructive knowledge, a divergence from entirely subjective towards more objective conceptions of the *mens rea* of knowledge, belief and suspicion is something that the law already takes, and which the hybrid approach to these concepts readily incorporates.

### *10.3.2. Testing Hybrid Knowledge, Belief, and Suspicion in Jurisprudence*

#### *10.3.2.1. R v Saik*

The House of Lords case of *R v Saik*[219] substantively concerned the interpretation of legislation relating to the inchoate offence of conspiracy and,[220] more specifically, the precise *mens rea* required for a conspiracy to commit money laundering. The case confirms some important points regarding knowledge; crucially, that knowledge consists of a true belief,[221] and that it is not possible to 'know that something is A when in fact it is B.'[222] Concerning the substantive issues in the case, however, the full reasoning of the

[216] Stephen Shute, 'Knowledge and belief in the criminal law' in Shute S. and Simester A. (eds.), *Criminal Law Theory: Doctrines of the General Part* (Oxford University Press 2002), 199 – 200.
[217] *R v Russell* (1985) 81 Cr App R 315.
[218] Sullivan (2002), 211.
[219] *R v Saik* [2006] UKHL 18.
[220] Criminal Law Act 1977, s. 1.
[221] *Saik* [2006], [26].
[222] *Ibid*., [70]; citing *Montila* [2004], [27].

Court need not be rehearsed here save for the Court's conclusion that, whereas the substantive offence of money laundering could be committed with the *mens rea* of knowledge or reasonable grounds of suspicion that money was the proceeds of crime, the inchoate offence of conspiracy to commit money laundering could only be committed with knowledge. This view of conspiracy is, however, departed from in the dissenting judgment of Baroness Hale, was reached uneasily in the judgment of Lord Nicholls,[223] and has been recommended for reform by the Law Commission.[224] As the facts of the case make clear, below, the defendant can have been in little doubt that he was in reality engaged in money laundering, and the law of conspiracy likely needs reforming to require either *mens rea* of knowledge or recklessness as to facts relevant to the conspiracy.[225]

The first defendant operated a modest bureau de change making annual profits of around £8,000; however, a second defendant exchanged money over a short period of time amounting to some $8 million, and the two were covertly filmed a number of times talking by the second defendant's car in which large bags of money had also been seen. The first defendant pleaded guilty to conspiracy to launder money, but on the basis that he only admitted to suspecting that the money was the proceeds of crime and that he never had explicit knowledge of the same. His conviction was subsequently quashed on the technicality that, as stated above, the crime of conspiracy could not be committed with suspicion alone; however, there is little doubt that the defendant had in fact committed money laundering which can be satisfied with the lesser *mens rea* of reasonable cause for suspicion. Equally, if conspiracy were opened up to include both knowledge and recklessness as advocated by the Law Commission, a just result would have been reached and the defendant's conviction upheld.

The same follows applying the hybrid tests. Starting with knowledge, under the first limb it was indeed certain that the money being converted in *Saik* was the proceeds of crime. However, applying the second limb of the test to the information available, it cannot necessarily be said that it is reasonable to expect anybody in the defendant's position to

[223] *Saik* [2006], [33].
[224] Law Commission, *Conspiracy and Attempts: A Consultation Paper* (Law Com No 183, 2007), 54 – 71.
[225] Jeremy Horder, 'Reforming the auxiliary part of the criminal law' (2007) 10 *Archbold News* 6, 8 – 9; see also Graham Virgo, 'Laundering conspiracy' (206) 65(3) *Cambridge Law Journal* 482.

appreciate that *certainty*. Whilst the defendant admitted to suspecting the criminal origins of the funds, and may have even suspected this to a very high degree, there is nothing in the evidence from which it is reasonable to conclude that *anybody* in the same circumstances would *know* as a *certainty* that the money was the proceeds of criminality. Of course, it might also be readily argued that the defendant was wilfully blind to the criminal origins of the money which itself may amount to knowledge; however, as the case proceeded on a guilty plea and never progressed through trial, this argument was never tested in court.

Applying a hybrid approach to wilful blindness (which requires suspicion plus the deliberate omission to make inquiries), it may first be stated objectively that there existed the conjecture that funds were from criminal origins. Second, it is reasonable to expect that anybody in the defendant's circumstances would appreciate this conjecture, not least from the fact of such a small and relatively unprofitable bureau de change suddenly exchanging millions of dollars; (and, indeed, the defendant admitted such suspicion). Third, the defendant patently failed to make any further inquiries whatsoever, such that it may reasonably be said that he was wilfully blind and, therefore, guilty of the conspiracy to launder money.

For completeness, it may also be seen that the defendant would be convicted of conspiracy if the *mens rea* requirement was expanded to include recklessness, as proposed by the Law Commission. Applying the first limb of the hybrid test for recklessness, there was clearly an unreasonable risk that the money received by the defendant came from criminal activities, again, not least due to large sums of money being delivered in bags of cash to an otherwise minor bureau de change. Further, applying the second limb, it is again reasonable to expect that anybody in the defendant's circumstances would appreciate that risk, both from the general knowledge of running such a business as a bureau de change and the specific indication from a sudden increase in business by the order of millions. Thus, again, the just result is achieved through a hybrid approach to knowledge (wilful blindness) and recklessness.

#### *10.3.2.2. R v Bello*

The facts of *R v Bello*[226] were briefly presented above, concerning a defendant convicted for knowingly remaining within the UK beyond the time permitted under his conditions of entry.[227] The defendant asserted that he had not *knowingly* remained in the UK beyond the permitted time because he had received the news of the death of his mother shortly before that time, which had devastated him, made it impossible for him to manage his business affairs, and had destroyed his memory and consumed his thoughts. The trial judge was heavily rebuked by the Court of Appeal because he had determined that no legal defence was offered and declined to leave the matter to the jury, whereas there 'might well [have been] questions of fact and degree as to the precise state of the defendant's mind which might [have arisen] for consideration.'[228] Nonetheless, the defendant's conviction was upheld on appeal.

The Court of Appeal considered that to possess knowledge of a fact was not the same as to be immediately thinking about it,[229] and it was both unnecessary and impractical to prove that a particular fact was forefront in a person's mind at a given time. Rather:

> '[A] man cannot plead that he did something unknowingly if he had the capacity for reviving the recollection of that event from his memory. A man can do an act knowingly even though at the moment when he does it the relevant fact is not actually in his mind. If he has the capacity to restore that fact to his mind, then on the face of it we would have thought the requirement of "knowingly" is satisfied.'[230]

This is a clear expression of Duff's "tacit" or Devlin's "second degree" knowledge. Although the defendant's defence had not been left to the jury in this case, the Court of Appeal considered that the outcome would have been the same regardless. Whilst no doubt distraught by his mother's passing, the defendant had been capable of maintaining

[226] *R v Bello* (1978) 67 Cr App R 288.
[227] Immigration Act 1971, s. 24(1)(b)(i).
[228] *Bello* (1978), 290.
[229] *Ibid*; citing Glanville Williams, *Criminal Law: The General Part* (2nd ed. Stevens & Sons Ltd. 1961), 170.
[230] *Ibid*.

a relatively normal life, including attending a polytechnic, such that the knowledge of his leave having expired would have been readily recallable.

For much the same reasoning, it is highly likely that the defendant in *Bello* would be similarly convicted under the hybrid approach to knowledge. Applying the first limb of the test, it was indeed certain that the defendant had outstayed his leave by the relevant date. Regarding the second limb – and as the Court of Appeal rightly determined – the defendant's state of mind following the news of his mother's passing is indeed a relevant circumstance in considering whether it is reasonable to expect anybody in the same circumstances to appreciate the certainty of the fact of having outstayed their leave to remain in the country. Whilst the effects of grief may well be appreciated, ordinary grief alone is not generally regarded as so entirely obliterating a person's access to knowledge or, indeed, legal responsibility for their actions. Knowing well in advance the date upon which the defendant was required to leave the country and having such ready means of reminder as keeping a record in a diary or calendar, it is reasonable to expect anybody experiencing ordinary grief to nonetheless appreciate the certainty that the date for their leave to remain has passed.

The defendant would need to demonstrate something more than ordinary grief – to the extent that it is significantly impacting on their capacities to appreciate the consequences of their actions, for reasons responsiveness or for ordinary self-control – before the law accepts that there is a valid defence. Following the evidence of the defendant's ability to continue attending his polytechnic and other daily activities, it cannot reasonably be concluded that his ordinary grief had elevated to something more severe – such as clinical depression, for example – that would impact across his aforementioned capacities and potentially provide the basis for a legal defence. Thus, the hybrid approach to knowledge demonstrably encompasses cases of tacit knowledge, as with the current law.

#### *10.3.2.3. R v Russell*

The facts of *R v Russell*[231] were similarly presented above and, briefly, concerned a defendant who had taped a knife inside a compartment under the dashboard of his car and hidden a cosh (a rubber hose containing metal at one end) under the seat. Upon prosecution for possession of an offensive weapon in public,[232] the jury found the defendant not guilty with respect to the knife, accepting his explanation that he used it as an ordinary tool and that it had been taped at the time to prevent it moving around when the vehicle was driving. Regarding the cosh, however, the defendant asserted that he had placed it in the vehicle some time ago intending to use the metal and, at the time of his arrest, he had forgotten completely about its existence within his car. The defendant was convicted in relation to the cosh and appealed on the grounds that, having entirely forgotten that the cosh was in his vehicle, he did not possess the requisite knowledge that he was in possession of an offensive weapon in public. This argument was accepted, the Court of Appeal finding that:

> 'It would be wrong in our judgment to hold that a man knowingly has a weapon with him if his forgetfulness of its existence or presence in his car is so complete as to amount to ignorance that it is there at all. This is not a defence which juries would in the ordinary way be very likely to accept, but if it is raised it should be left to them for their discretion.'[233]

The case of *Russell* would therefore be categorised as one of latent knowledge which, it seems, is insufficient for the *mens rea* of knowledge. However, the justice of the decision in comparison to *Bello* is perhaps difficult to appreciate; in *Bello*, the defendant cited what were potentially very persuasive explanations as to why he might have forgotten the date for the expiry of his leave on account of grieving for the loss of his mother. Although the law may not wish to make a legal defence out of grief, it nonetheless offers some explanation for the defendant's actions. Contrast this with *Russell*, however, where the defendant's only defence was that he had simply forgotten the fact of having placed an

[231] *Russell* (1985).
[232] Prevention of Crime Act 1953, s. 1.
[233] *Russell* (1985), 319.

offensive weapon in his vehicle; it seems, if anything, that Russell is the more culpable party.

Indeed, the decision in *Russell* has been treated with significant scepticism. The Court of Appeal in *R v Martindale*[234] later considered the "forgetfulness" argument to be 'fallacious' and rightly highlighted that were such a defence to be allowed, 'a man with poor memory would be acquitted, [whilst] he with the good memory would be convicted.'[235] With regards to *Russell* specifically, the Court of Appeal in *Martindale* opined that the earlier Court had not been properly referred to previous authority which, again, clearly precludes a forgetfulness defence in relation to knowingly being in possession of certain articles.[236] Similarly in *R v McCalla*[237] – a case following remarkably similar facts to *Russell* involving a defence of forgetfulness regarding a cosh hidden in the defendant's car – the Court of Appeal disavowed the mere forgetfulness defence.[238] Notably, however, the Court stresses that this has no bearing on cases where the defendant had no knowledge of possession of a forbidden item in the first place, for example, when some article is hidden upon their person or placed in their vehicle by another party.[239]

Read together, the current authorities appear to suggest that both tacit and latent knowledge may suffice for the *mens rea* of knowledge, if there is even any great distinction between the two. The Court of Appeal stated in *R v Buswell* (and has since approved in *Martindale* and *McCalla*):

> '[I]f you have got [some forbidden article] in your custody and you put it in some safe place, and then forget you have got it, and discover a year or two later, when you happen to look in that particular receptacle that it is still there, it seems to this court idle to suggest that during those two years

[234] *R v Martindale* (1987) 84 Cr App R 31.
[235] *Ibid*., 33.
[236] *Ibid*., 33 – 34; citing *R v Buswell* [1972] 1 WLR 64, 67.
[237] *R v McCalla* (1988) 87 Cr App R 372.
[238] *Ibid*., 378.
[239] *Ibid*; confirming *R v Cugullere* (1961) 45 Cr App R 108.

it has not been in your possession. It has been there under your hand and control.'[240]

It is readily submitted that the hybrid approach to knowledge would follow *Buswell*, *Martindale* and *McCalla*, and not *Russell*. Applying the first limb to the facts of *Russell*, it was indeed certain that the cosh was within the defendant's vehicle – the item had been found in its hiding place by the police. Turning to the second limb, the only relevant circumstance presented by the defence for consideration was the fact that he had forgotten placing the cosh in his car. The relevant question becomes, therefore, *is it reasonable to expect anybody who places a weapon in their car to appreciate the certainty that it is there, notwithstanding that they have later forgotten that fact?* For the reasons set out in the jurisprudence considered, the answer to this question must be affirmative; it is untenable that by claiming mere forgetfulness defendants should be able to avoid all liability for any offences requiring the *mens rea* of knowledge. Of course, defendants may suffer from the effects of some further illness or condition that induces their forgetfulness, in which circumstances a positive defence may follow; but the law surely cannot permit the ordinary, everyday forgetfulness suffered by all reasonable people to provide a defence on its own without rendering the *mens rea* of knowledge practically meaningless.

#### 10.3.2.4. *Westminster City Council v Croyalgrange Ltd.*

The previously considered cases discuss examples of explicit, tacit and latent knowledge where, contrary to popular statements in academic commentary, all three appear to be sufficient to satisfy the *mens rea* of knowledge, both in existing jurisprudence and the proposed hybrid approach. *Westminster City Council v Croyalgrange Ltd.*[241] is a leading House of Lords decision concerning the further issue of wilful blindness or Devlin's second degree of knowledge. The defendant company and individual company director were landlords of a property and were charged with knowingly causing or permitting premises to be used as a sex establishment contrary to the relevant licensing requirements, in relation to the use of that property by their tenant as a sex shop.[242] The substantive

[240] *Buswell* [1972], 67.
[241] *Westminster City Council v Croyalgrange Ltd.* (1986) 83 Cr App R 155.
[242] Local Government (Miscellaneous Provisions) Act 1982, Sch. 3, para. 20(1)(a).

question on appeal was whether or not it had to be proven that the defendant knew both that premises were being operated as a sex establishment *and* that this was not being done in compliance with licensing requirements, which was answered in the affirmative. However, considering whether this placed too onerous a burden upon the prosecution, the House of Lords affirmed that:

> '[I]t is always open to the tribunal of fact, when knowledge on the part of a defendant is required to be proved, to base a finding of knowledge on evidence that the defendant had deliberately shut his eyes to the obvious or refrained from an inquiry because he suspected the truth but did not want to have his suspicion confirmed.'[243]

The defendant's acquittal was upheld because the prosecution had failed to prove knowledge of both the use of the premises and the absence of a licence.[244] Had both of these questions been tested, however, it is readily appreciable that the defendant landlord may have been convicted. As the House of Lords noted, whereas the defendant tenant *actually* using premises as a sex establishment is in the best position to know whether the relevant licensing requirements having been complied with, a defendant landlord who is permitting such use of their premises 'likewise has the means of knowledge readily available to him.'[245] It being established that the defendant in *Croyalgrange* did indeed know of how their premises were being used, it can be appreciated why that defendant may be found to have been wilfully blind with regards to the licensing of that establishment if, suspecting that it had not been properly licensed, they failed to make further, readily available inquiries, such as by asking their tenant or checking with the relevant licensing authority.

In the present thesis, it is submitted that wilful blindness is readily incorporated into the hybrid approach to knowledge through the second limb of the test in particular. Recalling that knowledge must be factually true in law, the objective definition of knowledge stands

---

[243] *Croyalgrange* (1986), 164.
[244] See further Lynne Knapman, 'Permitting use of premises as unlicensed sex establishment' (1986) (Oct) *Criminal Law Review* 693.
[245] *Croyalgrange* (1986), 163.

in *Croyalgrange* as the certainty that the tenant was operating an unlicensed sex establishment. When turning to the second limb, it may readily be argued that it is reasonable to expect anybody to appreciate the aforementioned certainty if they suspect it to be the case and decline to make readily available inquiries in order to avoid confirming that suspicion. Put differently, it may readily be argued that it is not reasonable for anybody to claim that they did not appreciate a particular fact because, although they had suspected that fact to be so, they had declined to take readily available steps to confirm or deny that suspicion. Applying this to the present case, the defendant had indeed made inquiries from his tenant regarding their licensing situation and had been assured that the relevant license was obtained. If this evidence is accepted, then it may be argued that the second limb of the test for hybrid knowledge was not satisfied; it is reasonable to expect anybody who has received similar assurances to no longer appreciate the *certainty* that a particular circumstance exists contrary to those assurances.

The fact that wilful blindness so fits into the hybrid approach to knowledge may be further demonstrated by breaking down wilful blindness itself. If wilful blindness consists of a suspicion combined with the declination to make readily available inquiries then, so broken down within the present thesis, wilful blindness thus incorporates the hybrid definition of suspicion – the *conjecture that a particular circumstance exists* plus the reasonableness test – along with the failure to make further, readily available inquiries. Concerning the first limb of the test for suspicion, the peculiar facts of this case readily raise the objective conjecture that the tenant had not obtained a license for their sex establishment. In particular, the licensing requirement had only come into effect a couple of years after the property was originally let, whilst the tenancy agreement between the relevant parties was renewed every six months and had, at times, been amended to reflect changes in the law.[246] As the premises had never previously obtained a license to operate as a sex establishment, which was a fact admitted by the defendant, there obviously existed the conjecture – the suspicion, uncertainty or potential – that the premises might not have become so licensed by the necessary date.

[246] *Ibid.*, 156 – 159.

With regards to the second limb of the test, once again the defendant's evidence included the assertion that he had enquired from his tenant regarding the licensing of the premises and was assured that the relevant requirements had been complied with. The similar argument follows that this may be sufficient to conclude that it would not be reasonable to expect the defendant to appreciate the continuing conjecture that premises were being used as a sex establishment without a license. It is the point of reasonableness that is crucial to this argument – what steps does the law *reasonably* expect a landlord to take in order to satisfy themselves that their tenant possesses the necessary licensing and, specifically, is a simple enquiry to that tenant sufficient? If so – which appears to be the suggestion from *Croyalgrange*[247] – then it is not reasonable to expect anybody in such circumstances to continue appreciating the very conjecture that they have discounted. Then again, it might be argued that obtaining a mere verbal assurance from the tenant are not necessarily sufficient grounds upon which any reasonable defendant can be excused from appreciating the conjecture that that tenant is operating an unlicensed business. Once again, this ultimately represents the dividing line upon which it is for the jury to determine what are the bounds of reasonable and unreasonable behaviour.

On balance, the first approach is arguably preferable; suspicion is a particularly low threshold of *mens rea* to satisfy, and it should not, therefore, require an especially high level of inquiry or investigation in order for it to be reasonable to expect that anybody would cease appreciating a particular conjecture. This is to say that if, on the facts of the case (including circumstances subjective to the defendant such as information known or available to him), it may be said that any other reasonable defendant in the same circumstances would not be concerned with a particular suspicion or conjecture, then the hybrid test for suspicion is not satisfied. On the facts presented in *Croyalgrange*, the hybrid test would likely not be satisfied, and so the defendant would not meet the requirements for wilful blindness. Supposing that hybrid suspicion was found, however, then the second consideration under wilful blindness is whether or not the defendant has failed to make readily available inquiries so as not to confirm their suspicions.

[247] *Ibid.*, 163 – 164.

This, again, draws the question back to what information was readily obtainable by a defendant or what inquiries could easily or obviously have been made.[248] In the present case, a readily available means of inquiry was asking the tenant whether they had obtained the necessary licenses in accordance with their legal obligations. This the defendant duly did and was assured that the license had been obtained; the defendant had arguably not failed to make readily available inquiries, therefore, and is not wilfully blind following the hybrid approach. Such a contextualised approach to wilful blindness is similarly discussed by Wasik and Thompson who highlight that the point of fault under wilful blindness is the deliberate closing of one's eyes so as not to have suspicions confirmed. From this perspective, it is relevant to consider the entire context of the defendant's situation and, for example, if it would be 'almost impossible, or perhaps even inappropriate, to take steps to convert his suspicion into knowledge, then [the defendant] is not *wilfully* blind.'[249]

#### 10.3.2.5. *R v Griffiths*

The case of *R v Griffiths*[250] offers a further example of wilful blindness within the context of handling stolen goods which may be committed where, *inter alia*, the defendant knows or believes that the goods in question are in fact stolen.[251] The defendant – who was also convicted for an unrelated burglary and had a long history of stealing, burglary and theft – was convicted of handling stolen good in relation to some candlesticks found in his possession that had been stolen from a church a few days prior to his arrest. The defendant had attempted to sell the candlesticks at a market on the afternoon of his arrest and, by his own admission, had lied to two dealers about how he came to possess the items. The defendant also gave mixed accounts to the police, first stating that he had purchased the candlesticks from a dealer that same afternoon, and later stating that he had bought them from a man on the high street (who the defendant was curiously unable to describe). When asked by the police whether he had inquired where the candlesticks came from, the defendant replied saying "you don't ask questions like that, do you?" When it was

[248] Per Sullivan (2002), 214, wilful blindness does not require that a defendant fails to make extensive or even reasonable inquiries, but only those such as are readily available.
[249] Wasik and Thompson (1981), 336.
[250] *R v Griffiths* (1974) 60 Cr App R 14.
[251] Theft Act 1968, s. 22.

suggested to him that he must have realised they were stolen, he further answered "yes, I suppose so", albeit he later denied this latter statement in his written evidence. He further stated in evidence at trial that he "might have had suspicions, but the suspicions were not related to any criminal offence."[252]

The defendant appealed on two grounds, the second of which is of particular interest concerning the status of wilful blindness in relation to knowledge. Whereas the judge at trial had directed the jury that wilful blindness equated in law to a third culpable state alongside knowledge or belief which may be satisfied for the offence of handling stolen goods, the Court of Appeal stressed that wilful blindness did not amount to knowledge as a matter of law, but entitled a jury to infer knowledge.[253] Specifically, the Court guided:

> 'To direct the jury that, in common sense and in law, they may find that the defendant knew or believed the goods to be stolen because he deliberately closed his eyes to the circumstances is a perfectly proper direction.'[254]

Although the Court of Appeal considered that the trial judge had potentially misdirected the jury as to the status of wilful blindness specifically, it was nonetheless satisfied that the judge's direction overall was suitable. Moreover, even had the Court determined that the misdirection was too overbearing, they would have had no hesitation in maintaining the defendant's conviction, as 'the evidence was so overwhelming and no reasonable jury on this evidence could have arrived at a conclusion other than that the [defendant] believed the goods were stolen.'[255]

Approaching wilful blindness in this case first through the context of hybrid knowledge, the first limb of the test is clearly satisfied as it was certain that the candlesticks in the defendant's possession were indeed stolen. Concerning the second limb of the test, according to the police the defendant purportedly did accept the suspicion that goods

[252] *Griffiths* (1974), 14.
[253] See Wasik and Thompson (1981) for a robust defence of wilful blindness equating to knowledge or belief as a matter of law and not fact.
[254] *Griffiths* (1975), 18.
[255] *Ibid*., 18 – 19.

might have been stolen and, furthermore, that he had not made further inquiries from the seller of the candlesticks because it was not the "done thing". It is also relevant that the defendant had admittedly lied when attempting to sell the candlesticks which, it may be suggested, reveals his suspicions as to their unlawful provenance. Given all of the above, it is readily arguable that it is reasonable to expect anybody in the same circumstances to appreciate the fact that the goods were stolen; it would be *unreasonable* for the defendant to claim that he suspected the goods to be stolen but did not wish to ask the man selling them on the high street because he felt inappropriate in so doing or, even worse, did not care to hear the answer. This perspective on wilful blindness fits with the Court of Appeal's designation of the same entitling the inference of knowledge whilst not necessarily equating to the same thing. In a similar vein, placing wilful blindness within the second limb of hybrid knowledge is not to suggest that wilful blindness is equivalent to a person *actually* possessing knowledge of a thing, but that it is *unreasonable* for defendants to deny such knowledge in circumstances where they have been wilfully blind.

As above in relation to *Croyalgrange*, it is also possible to break wilful blindness down into its constituent components and consider how these fit within the hybrid approach to *mens rea*. First, clearly there existed the conjecture in *Griffiths* that the property was stolen, whether taken from the defendant's own admission of suspicion or the circumstances of his having been approached by a man on the high street to purchase candlesticks at half-price. Second, there is nothing in the facts to suggest that it is not reasonable to expect anybody in the same circumstances to appreciate this conjecture; the defendant offered no reasons why he did not or could not appreciate the suspicious nature of the candlesticks. Finally, the defendant had demonstrably failed to take readily available steps to allay any suspicion – namely by asking the man on the high street of the provenance of the goods – because, by his own admission, "you don't ask questions like that, do you?" Thus, whether considered broadly through the reasonableness limb of hybrid knowledge or broken down and considered through the application of hybrid suspicion plus the declination to make readily available inquiries, the concept of wilful blindness is again readily incorporated into the hybrid approach to *mens rea*.

#### *10.3.2.6. Atwal v Massey*

Guidance on constructive knowledge is provided in the case of *Atwal v Massey*[256] which, again, concerned the offense of handling stolen goods. Following somewhat curious facts, one man stole a kettle and left it outside a gate by a roadside junction for the defendant to collect, who later paid the thief. No evidence is reported concerning the defendant's account of events or knowledge or beliefs regarding the thief and kettle, beyond the facts as stated. The defendant was convicted in the Magistrate's Court on the basis that, 'although there were discrepancies in the evidence, the defendant from the circumstances under which he had collected the kettle ought to have known that it was stolen.'[257] This introduced the concept of constructive knowledge – *i.e.,* that which a person *ought* to have known or what the reasonable person would have known in the same circumstances.

Upon appeal, however, the Divisional Court, led by the Lord Chief Justice, resolutely concluded that this was a misdirection; in order to establish the requisite knowledge or belief for the offense of handling stolen goods, the Court considered it insufficient to demonstrate that the goods were 'received in circumstances which would have put a reasonable man on his enquiry.'[258] The defendant was consequently successful in his appeal and acquitted of the substantive offence. It must be highlighted, however, that whereas this precludes the inclusion of constructive knowledge within the *mens rea* of knowledge generally, as indicated earlier in this section, Parliament has introduced a number of offences which can explicitly be satisfied with constructive knowledge.

A similar outcome can be reached through the application of hybrid belief. First (and following only from the available facts as presented), the Divisional Court was correct in stating that the 'whole case reeked with suspicion' and, indeed, the circumstances in which the defendant received the stolen kettle were peculiar. However, it is submitted that the mere fact of the defendant collecting the item from a prearranged location – outside a gate, which could in principle have been their own or that of a legitimate seller – does not support the *conviction* that the property was stolen. In other words, the

[256] *Atwal v Massey* (1972) 56 Cr App R 6.
[257] *Ibid.,* 7.
[258] *Ibid.,* 8.

circumstances could also support the reasonable conclusion that the kettle was being purchased from a neighbour or associate who, given the low-value and informal nature of the transaction, had arranged to drop off the item at a given location. Thus, the prosecution's case falters at the first hurdle; whilst the presented circumstances support the *conjecture* that the kettle was stolen, they do not necessarily support the same *conviction*.

#### *10.3.2.7. R v Hall*

*R v Hall*[259] is another case concerning the offense of handling stolen goods but, rather than looking at the concept of wilful blindness, focuses instead on the distinction between knowledge and belief and the latter *mens rea* in particular. The case was cited in section 10.3 of this thesis, above, for providing a definition of belief as something short of knowledge but for which there is no other reasonable conclusion. Two people – (who were convicted of burglary in relation to the same case) – met the defendant outside a block of flats, passed him one of three suitcases that they were carrying, and then proceeded into one of the flats. They were observed by the police who knocked at the door of the flat ten minutes later. Inside, the police found a blanket on the floor concealing a substantial number of items of silver cutlery, further silver cutlery in a kitchen drawer, some paintings on the floor, and several ornaments 'tidily arranged on shelves in the room and on the television set.'[260]

In so far as it relates to the issue of knowledge or belief, the defendant's evidence from his first police interview was that the two other persons had attempted to hide some of the property around the flat when the police knocked at the door. They purportedly told him that the property came from a house clearance, but the defendant realised that this was 'no house clearance in the accepted sense of that term.' He asserted that he knew the property was stolen when the men started to hide the property upon the arrival of the police. In his second police interview, however, the defendant admitted to being told by the two men that they had committed a burglary and that he knew they were bringing

[259] *R v Hall* (1985) 81 Cr App R 260.
[260] *Ibid.*, 261.

stolen goods to him; however, he asserted in evidence at trial that this had been a sarcastic comment and not intended as an admission. At trial, the defendant repeated that he had concluded from the way some of the property was wrapped that it was 'not from a house clearance in the ordinary sense of that term, because it was clear that these people were not professionals in that particular sphere.'[261] Finally, whilst giving evidence at trial *for the defendant*, one of the burglars repeated twice that the defendant had stated immediately upon seeing the items, '"those are not from a house clearance, I don't want to have anything to do with them."'[262]

The defendant was duly convicted of handling stolen goods, which the Court of Appeal had little difficulty in upholding. The defendant appealed on the grounds, *inter alia*, that the trial judge had misdirected the jury regarding knowledge and belief, and conflated belief with suspicion; however, the Court found that the trial judge's directions to the jury were impeccable. On the matter of *mens rea* specifically, the Court commented:

> 'A man may be said to know that goods are stolen when he is told by someone with first hand knowledge (someone such as the thief or the burglar) that such is the case. Belief, of course, is something short of knowledge. It may be said to be the state of mind of a person who says to himself: "I cannot say I know for certain that these goods are stolen, but there can be no other reasonable conclusion in the light of all the circumstances, in the light of all that I have heard and seen. Either of those two states of mind is enough to satisfy the words of the statute. The second is enough (that is, belief) even if the defendant says to himself: "Despite all that I have seen and all that I have heard, I refuse to believe what my brain tells me is obvious". What is not enough, of course, is mere suspicion. "I suspect that these goods may be stolen, but it may be on the other hand that they are not".'[263]

[261] *Ibid.*, 262.
[262] *Ibid.*
[263] *Ibid.*, 264.

The Court of Appeal further specifically approved of the trial judge's directions:

> 'You know what believing means because we all believe things every day. We look at all the circumstances and we make up our minds about something, we come to a belief about them having looked at all the circumstances of the case and we say yes, everything points in that direction and I believe that such and such is a fact… The law says you cannot simply say it is perfectly obvious but I am not going to believe it. In other words, you cannot shut your eyes to the plain and obvious.'[264]

Whilst the Court maintains that knowledge and belief are entirely subjective concepts, hints of objectivism can be appreciated in this judgment, in particular where belief is described as consisting of "no other *reasonable* conclusion" and where a defendant is precluded from disbelieving something that is "plain and obvious." That notwithstanding, on the evidence provided it was patently clear that the defendant in *Hill* at least believed the goods to be stolen, not least drawing from his inconsistent accounts and the evidence from the burglar asserting that the defendant had recognised the unlawful origins of the items upon first seeing them.

It is equally appreciable that the defendant would be similarly convicted under the hybrid approach to belief. Starting with the first limb, the question is whether there objectively existed the *conviction* that the goods in question were stolen – *i.e.,* on the facts presented, was there any reasonable conclusion other than the goods were stolen. If the second police interview is accepted where the defendant purportedly admitted being told by the burglars that they were bringing stolen items, this conviction is clearly made out. Even on the remaining evidence, however, it is readily arguable that such an objective conviction can be found: from the mis-matched collection of silverware, art, ornaments; to their inappropriate wrapping and delivery in three suitcases; the incongruence between these facts and the burglars' assertion to be professional antiques dealers; and their hiding the goods around the flat when the police knocked at the door – all the circumstances readily support the objective conviction that the goods were stolen. Similarly, on the second limb

[264] *Ibid.*, 264 – 265.

of the test, the defendant offers nothing in particular (beyond his bare denial of knowledge or belief) to suggest that it would not be reasonable to expect anybody in the same circumstances to appreciate the conviction that the goods were stolen.

It may be useful to contrast the finding of hybrid belief in *Hall* with the absence of a similar finding in *Atwal*, above, owing to the notably different facts. Whilst the manner of the sale and delivery of the kettle in *Atwal* was appreciably *suspicious* and out of the norm, it could equally be explained in the context of some private sale of an item between neighbours, one of whom left the item for another to pick up from a prearranged location. That is to say, there is another reasonable explanation in *Atwal* such that, whilst suspicion may indeed be present, belief is not. In *Hall*, however, there can be no other such reasonable explanation of the totality of the circumstances other than the *conviction* – the conclusion or belief, but not necessarily the knowledge or certainty – that the goods were stolen. Observing this contrast between *Atwal* and *Hall* offers some further insight into the application of the first limb of the hybrid test for belief in particular.

#### *10.3.2.8. R v Da Silva*

The case of *R v Da Silva*[265] is one of the few authorities in criminal law discussing the *mens rea* of suspicion in any particular detail. The defendant was charged and convicted of entering into an arrangement to conceal or transfer the proceeds of another's criminal conduct, knowing or suspecting that that other person had, in fact, been engaged in criminal conduct or benefited therefrom.[266] The other person in question was the defendant's husband and co-accused, who had been convicted for the substantive offence of obtaining money by deception when, in his position as the manager of a coffee-bar, he had submitted false employee timesheets and directed payments towards two bank accounts held and operated by the defendant. The defendant herself had been charged with these substantive offences but was found not guilty. This acquittal thereby excluded knowledge as a possible *mens rea* for the offence of concealing or transferring the proceeds of another's criminal conduct – had she had the requisite knowledge, the

[265] *R v Da Silva* [2006] EWCA Crim 1654.
[266] Criminal Justice Act 1988, s. 93A(1)(a).

defendant would have been guilty of the substantive offence – leaving only the issue of the *mens rea* of suspicion.[267]

In the absence of specific authority from the criminal courts, the Court of Appeal in *Da Silva* turned to the civil courts for inspiration, first accepting a 'state of conjecture or surmise' as providing a 'general indication of the general meaning of "suspicion".'[268] The Court proceeded to offer:

> '[T]he essential element in the word "suspect" and its affiliates, in this context, is that the defendant must think that there is a possibility, which is more than fanciful, that the relevant fact exists. A vague feeling of unease would not suffice. But the statute does not require the suspicion to be "clear" or "firmly grounded and targeted on specific facts", or based upon "reasonable grounds".'[269]

From this, it may be appreciated that suspicion sets a relatively low *mens rea* threshold, albeit is still something more than *de minimis*. Whilst the trial judge had provided a misdirection on suspicion, in the event the Court of Appeal considered that there was no doubt as to the safety of the conviction. The defendant had declined to answer during police interview, and provided no evidence other than the assertion that her husband had instructed her to make a bank account available for employees who had none of their own. This explanation had already been plainly rejected, however, in the conviction of the defendant's husband.

Considering *Da Silva* through the hybrid approach, the circumstances described in the available facts readily support the conjecture – the possibility or suspicion – that the funds in question came from the defendant's husband's criminal activities. It is alone peculiar that a company running a chain of coffee bars would instruct a branch manager to open

---

[267] *Da Silva* [2006], [4].
[268] *Ibid.*, [13] – [14]; citing *Hussien* [1970], 948.
[269] *Ibid.*, [16]; see further David Ormerod, 'Proceeds of crime: Assisting another to retain benefit of criminal conduct knowing or suspecting other person to be engaged in criminal conduct' (2007) (Jan) *Criminal Law Review* 77, 78 – 79.

an account on behalf of employees in which to receive payments, whilst the otherwise unexplained appearance of ten payments of wages into the defendant's bank account readily supports the possibility or suspicion of the money's criminal origins. With regards to the second limb of the test, it is relevant that the defendant offered no explanations during her police interview, whilst the court at first instance clearly rejected her and her co-accused's explanation for receiving the money into her bank account. With nothing more proffered by the defendant, whose credibility is already in question following her rejected account of events, it is perfectly reasonable to expect that anybody in precisely the same circumstances would appreciate the conjecture that the monies they had received came from another's criminal activity.

#### 10.3.2.9. *R v Lane and Letts*

The case of *R v Lane and Letts*[270] concerned the appeal against a ruling of the trial judge during a preliminary hearing and, as such, concerns a matter that needed to be determined before the substantive trial had commenced. Consequently, there are scant facts and no evidence to be considered in this case, in which the defendants were charged with 'sending money overseas, or arranging to do so, when they knew or had reasonable cause to suspect that it would, or might, be used for the purpose of terrorism.'[271] The question on appeal was whether "reasonable cause to suspect" required that a defendant actually suspected a fact and had reasonable cause for such suspicion, or simply that there existed reasonable cause for suspicion, objectively assessed, on the information known to the defendant. As discussed in section 10.3.1, above, a great many offences may be committed *either* with a subjective element such as knowledge or belief, *or* with an objective element such as *reasonable cause* for belief or suspicion. As one of the most authoritative Supreme Court discussions on the issue, therefore, *Lane and Letts* is relevant for consideration.

Drawing from reasons of statutory interpretation – for example, including the fact that the Terrorism Act 2000 refers in various sections to both the *mens rea* of suspicion and

[270] *R v Lane and Letts* [2018] UKSC 36.
[271] *Ibid.*, [2]; Terrorism Act 2000, s. 17.

reasonable cause for suspicion[272] – the Supreme Court determined that it was not required for the prosecution to prove that the defendant *actually* suspected that funds were being used for the purpose of terrorism, only that there existed reasonable grounds upon which such a suspicion could be formed by the reasonable man.[273] Of greatest interest to the present thesis, however, is the Court's discussion of how the *mens rea* of reasonable cause for suspicion falls within the dichotomy between subjectivity and objectivity. In a paragraph that warrants repetition in full, the Supreme Court provided:

> 'In the present case it would be an error to suppose that the form of offence-creating words adopted by Parliament result in an offence of strict liability. It is certainly true that because objectively-assessed reasonable cause for suspicion is sufficient, an accused can commit this offence without knowledge or actual suspicion that the money might be used for terrorist purposes. But the accused's state of mind is not, as it is in offences which are truly of strict liability, irrelevant. The requirement that there exist objectively assessed cause for suspicion focuses attention on what information the accused had. As the Crown agreed before this court, on the information available to the accused, a reasonable person *would* (not might or could) suspect that the money might be used for terrorism. The state of mind of such a person is, whilst clearly less culpable than that of a person who knows that the money may be use for that purpose, not accurately described as in no way blameworthy. It was for Parliament to decide whether the gravity of the threat of terrorism justified attaching criminal responsibility to such a person, but it was clearly entitled to conclude that it did. It is normal, not unusual, for a single offence to be committed by persons exhibiting different levels of culpability. The difference in culpability can, absent other aggravating features of the case, be expected to be reflected in any sentence imposed if conviction results.'[274]

[272] *Ibid.*, [13] & [18] – [22].
[273] *Ibid.*, [6] & [25].
[274] *Ibid.*, [24] (original emphasis).

There are a number of important points to be raised from this paragraph. First, the Supreme Court provides what is perhaps that closest description in extant law of how the hybrid approach to *mens rea* proposed in this thesis is designed to operate. Where the Court explains the test as being what the reasonable person would suspect on the information available to the accused, it is encapsulating the second limb of the approach to hybrid *mens rea* in this thesis. Thus, the *mens rea* of reasonable cause for knowledge, belief or suspicion in the current law closely follows the hybrid approach, specifically in applying an objective test (for knowledge, belief or suspicion) to the defendant's subjective circumstances; *i.e.,* the "reasonable man" is placed into the defendant's shoes. The question always asked is, *is it reasonable to expect anybody* (*i.e.,* the reasonable man) *in the defendant's* (subjective) *circumstances to appreciate the nature and consequences of their actions* (as described by the objective definition for each particular type of *mens rea*).

Second, it should be noted that the House of Lords in *Saik* reached a different conclusion regarding the requirement for a defendant to possess a subjective suspicion for the *mens rea* of "reasonable cause for suspicion". However, it is equally important to note that the relevant comments in *Saik* were made *obiter dictum*,[275] were restricted to the particular money laundering offence under consideration,[276] and concerned a statutory offence that has since been superseded. In this regard, the conclusions of the Supreme Court in *Lane and Letts* are considerably more authoritative. As Thomas rightly notes, whereas the decision on this point in *Saik* was confined to the facts of that case, the decision in *Lane and Letts* 'offers clarity as to the operation of the phrase "reasonable grounds to suspect" generally.'[277]

Third, with respect generally to replacing subjective suspicion (belief and knowledge) with the hybrid objective / subjective approach, it is again reiterated that such a move ought not be as controversial as it may at first seem, given that the courts are already using an incredibly similar test as that being proposed for the *mens rea* of reasonable cause /

[275] *Ibid*., [17].
[276] *Ibid*; citing *Saik* [2006], [51] & [102].
[277] Mark Thomas, '"Reasonable cause to suspect": In the absence of knowledge and actual suspicion' (2018) 82(6) *Journal of Criminal Law* 423, 429.

grounds to suspect. The question then arises, under the hybrid approach, what distinguishes the *mens rea* of suspicion, belief or knowledge *simpliciter*, as contrasted against reasonable cause / grounds for the same? The distinction, it is proposed, is reflected in the first limb of the hybrid test where suspicion, belief or knowledge are defined objectively. Thus, where the *mens rea* of a particular offence requires one of the three aforementioned states *simpliciter*, the objective test follows as presented at the beginning of this section of the dissertation; for example, suspicion is defined objectively as the *conjecture that a particular circumstance exists*.

Where the requisite *mens rea* for an offence is reasonable cause / grounds for suspicion (belief or knowledge), however, the objective definition varies slightly to become the *reasonable conjecture that a particular circumstance exists*. Therefore, it must be demonstrated under the first limb that the circumstances of a given case support the conjecture that a particular circumstance existed *and* that that conjecture was objectively reasonable. Consequently, the second limb of the hybrid test becomes a rebuttable presumption in favour of the prosecution; the prosecution is not actively required to prove the second limb of the test once the first has been satisfied, whilst the defence may still adduce evidence to rebut the second limb. This is precisely because the requirement of reasonableness has been incorporated into the objective first limb of the test. It is recalled from section 9.1 of this thesis, above, that the hybrid approach to *mens rea* is intended to determine the boundaries of reasonable and unreasonable (as opposed to morally blameworthy) conduct. Thus, the second limb of the test asks whether it is reasonable to expect anybody – *i.e., the reasonable man* in the defendant's subjective circumstances – to appreciate the nature and consequences of their actions as they relate to the *mens rea* objectively defined in the first limb.

If a given certainty, conviction or conjecture has already been objectively determined to be reasonable, however, then it is automatically reasonable to expect that *the objective reasonable man* would appreciate that certainty, conviction or conjecture, *unless there are relevant circumstances which diminish or abrogate entirely their capacity to so appreciate the nature and conduct of their actions.* Put in the negative, it is not reasonable to expect that the *objective reasonable* man could not appreciate a certainty, conviction

or conjecture which is itself *objectively reasonable*, unless there are further circumstances (subjective to the particular defendant) which would so diminish the reasonable man's ability to appreciate the nature and consequences of their actions as it relates to that objective *mens rea*. Thus, the second limb of the test is presumed to be resolved in favour of the prosecution once the first limb is itself proven, unless the defendant can adduce facts or circumstances relevant to their capacities which rebut that presumption. In this way, "reasonable grounds" for suspicion (belief, *etc.*) should be easier to prove than mere suspicion under the hybrid formulation, just as objective reasonable grounds for suspicion is easier to prove than subjective suspicion under the current existing law. This rebuttable presumption is explored more fully in relation to the *mens rea* of negligence, discussed in section 10.5 of this thesis, below.

#### *10.3.2.10. R v B*

Having regard to the position reached by the Supreme Court in *Lane and Letts*, above, a final way that knowledge, belief or suspicion may be required for an offence is where it is the *absence* of knowledge, belief or suspicion that forms the requisite *mens rea*. The most obvious examples of this presentation of *mens rea* may be found within a number of sexual offences, such as the offence of rape which requires *inter alia* that the defendant *does not reasonably believe* that the victim consents,[278] or the offence of sexual activity with a child which requires *inter alia* that the defendant *does not reasonably believe* that the victim is aged 16 years or older.[279] It should be noted that, for the purposes of the offence of rape, whether or not a belief is reasonable 'is to be determined having regard to all the circumstances.'[280] The case of *R v B*[281] is interesting to consider in this respect as the defendant in question suffered from schizophrenia, inviting the discussion of how mental illness may (or may not) have a bearing on this *mens rea* and the *reasonableness* of any belief in particular. The defendant in this case was convicted of rape and common assault against his partner alongside minor criminal damage to her property; he had previous been convicted on counts of assault against the same victim, but they had since

[278] Sexual Offences Act 2003, s. 1(1).
[279] *Ibid.*, s. 9(1).
[280] *Ibid.*, s. 1(2).
[281] *R v B* [2013] EWCA Crim 3.

reconciled. Although, at trial for the earlier assaults, the available psychiatric evidence was that he did not suffer from any mental illness, considering the pattern of offences as a whole 'with hindsight, his illness was clearly developing.'[282]

The defendant in *R v B* did not give any evidence at trial and his case was effectively built upon answers that he had given to the police. With respect to the first count of common assault, the defendant was accused of remonstrating and spitting at his victim because she had passed some time with their (male) neighbour; the defendant told the police that no such incident took place, and that he had no objection to his partner talking with the neighbour. Regarding the second count of assault, the defendant was accused of forcing his victim to eat a bowlful of cold tinned peas containing crumbled leaves from a tree in the garden, first insisting that she ate the mixture, and then grabbing her finger and forcing her to do so. The defendant accepted in interview that he gave his partner this mixture, that she did not want it, and that he had told her to "have it" whilst standing by her, but he denied forcing her to eat it. With respect to the count of criminal damage, the prosecution alleged that the victim returned home from work one day to discover that the defendant had dismantled her doorstep; it was added that, on a previous occasion not subject to the charge at trial, the defendant had insisted on cutting down all the trees in the garden. The defendant asserted in interview that he had previously spoken with his partner about lowering the step and that she had agreed.

Concerning the first count of rape – alleged to have occurred the same evening as the second count of assault – the prosecution case was that the defendant told his partner that he wanted sex and insisted when she told him that she did not so want. In the bedroom, the defendant sprayed her genital area with a spray that purportedly smelled of bleach and caused a burning sensation, before having sex with her in a rough manner. The victim agreed that she had undressed herself prior to being sprayed but denied consent; 'in effect her evidence was that she objected but submitted in the face of his insistence.'[283] In interview, the defendant asserted that the spray was merely a scent but did not contain bleach, and that he never had intercourse with the victim without her consent. At trial, the

[282] *Ibid.,* [3].
[283] *Ibid.,* [12].

defendant's case was that his partner had indeed consented to having sex, and that he had told her that the spray would make her clean. The second count of rape is alleged to have taken place the same night when the defendant awoke wanting sex again; his partner said no and the defendant insisted, her evidence being that he was 'far stronger than she' and that he generally 'would not take no for an answer.'[284] The defendant again accepted that the sex had taken place but contended that his partner had consented.

The psychiatric evidence during the defendant's second trial is considerably more illuminating than that from his first. In brief, it was submitted that the defendant clearly suffered from paranoid schizophrenia or schizo-affective disorder, and had probably been suffering from this condition at the time of the alleged offences. Whilst a different psychiatrist had considered the defendant to be unfit to plead some six weeks after his arrest, the defendant had since been medicated and was found to be much better at the time of trial. As part of the defendant's mental illness, he believed that he had healing powers, and sexual healing powers in particular, with which he could cure cancer and epidemics. He also believed that he had a special connection with God and possessed solutions to the banking crisis and climate change. The defendant lacked insight into his illness when ill but gained greater insight when his health improved (for example, he no longer believed that had healing powers); however, his insight into his illness did remain impaired and he did not draw any connection between that and his behaviour towards his partner. Crucially, the psychiatric evidence submitted that the defendant had not been insane at the time of the alleged offences, and had retained the capacity to understand both what he was doing and the fact that it was wrong, although he had an impaired ability to interpret events normally, read signals or appreciate others' perspectives. With regards to the connection between his delusional beliefs and the offences:

> 'The acts of intercourse might have been motivated by his delusional beliefs that he had healing powers, including sexual healing powers… He might have believed that although she was saying no to sexual intercourse, it would still be good for her, and so he might have continued notwithstanding her response. Any such delusional beliefs did not,

[284] *Ibid.,* [15].

> however, extend to a belief that she was consenting; his illness was not relevant to his understanding whether she was consenting to sexual intercourse or saying no.'[285]

The Court of Appeal approved of the trial judge's "admirable" direction to the jury, beginning by explaining that the law does not permit a defence of mere mental illness as a matter of public policy. If such a defence existed then criminal conduct would be *carte blanche*; rather, the law recognises the influence of mental illness through such defences as insanity and diminished responsibility, and through sentencing. Concerning the reasonableness of any belief held by the defendant, the trial judge further directed that a delusional belief or a belief resulting from mental illness *cannot* be regarded as reasonable and, as such, the defendant's mental illness could not be taken into account when considering whether or not any belief he held was reasonable. Thus, the judge directed that it was necessary to ask whether:

> '[I]f you put the mental illness out of the question, were all the signs and signals such that someone who had been in a relationship with her all those years would have picked up on the signals and realised that she was not consenting, or were the signals such that someone would have, or might have, though "Yes she is consenting" and have carried on?'[286]

On approving this position, it was clear that the defendant's appeal must fail; his mental illness was not a relevant factor to the *reasonableness* of a belief. The Court of Appeal proceeded to refer again to the psychiatric evidence which had stated unequivocally that the defendant's mental illness had not affected his ability to understand that his partner was not consenting to sex, further supporting the safety of his conviction. The Court then reiterated that, even if the defendant's delusional beliefs could have led him to form the belief in his partner's consent, they could not have rendered that belief as being reasonable

[285] *Ibid*., [20].
[286] *Ibid*.

when, in fact, it was not; otherwise, 'then the more irrational the belief of the defendant the better would be its prospects of being held reasonable.'[287]

However, the Court confusingly added that their judgment did not exclude the possibility of cases arising where the 'personality or abilities of the defendant may be relevant to whether his positive belief in consent was reasonable', offering unusually low intelligence or some inability to recognise behavioural cues as potential examples of such a case.[288] This is particularly curious as the defendant in the present case had been described as having an impaired ability to read signals, and it is not explained why his schizophrenic disorder would be an irrelevant consideration whilst another's below average intelligence would not; surely, both of these conditions have the *potential* to significantly impact upon a person's understanding of another's behavioural signals and consent.

*

This lengthy exposition of *R v B* is necessary to appreciate both the criticisms of the decision and the way in which the hybrid approach to reasonable belief proposes to resolve those criticisms. For example, Laird contends that whereas the Court of Appeal in *R v B* requires a delusional belief in consent to be judged by objective standards of reasonableness without taking into consideration any mental disorder without which that belief would not have arisen, 'it is arguable that an individual who, through no fault of their own, is incapable of attaining the standard of a reasonable person is not blameworthy.'[289] This is very similar to the objections raised against purely objective *Caldwell* recklessness under which subjective characteristics such as a defendant's age or mental acuities could not be taken into consideration and which, in turn, led to demonstrable injustice in cases such as *Elliott v C*.

Indeed, the parallels with the *mens rea* of recklessness do not end there, as the more objective approach to reasonable belief adopted under the Sexual Offences Act 2003 and

[287] *Ibid*., [35].
[288] *Ibid*., [41].
[289] Karl Laird, 'Terrorism: *R v Lane (Sally)* Supreme Court: Lady Hale PSC, Lord Burnett CJ, Lords Hughes, Hodge and Mance: 11 July 2018' (2019) 2 *Criminal Law Review* 178, 180.

in *R v B* was intended to resolve similar problems that had arisen with purely subjective *Cunningham* recklessness, namely that under a purely subjective test, the more distorted or irrational a person's thinking is, the less likely it is for the test to be satisfied. Within the context of the offence of rape, this resulted in the legal defence of genuine (even if unreasonable) mistake as to consent,[290] which was pleaded with an unreasonable degree of success and became socially unacceptable to permit, not least in light of abysmal rates of successful prosecution for rape and other sexual offences.

This leads to a further criticism of *R v B* that, whilst the Court of Appeal cannot be faulted for their technical exercise in statutory construction, the interpretation arrived at arguably does not reflect the intentions of the government and Parliament in enacting the Sexual Offences Act 2003. As suggested above, a significant impetus for legal reform arose from the fact that a purely subjective test for belief in consent under the previous law resulted in the possibility that a defendant may claim that they genuinely believed in consent even though that belief was unreasonable from an objective viewpoint – the much-maligned unreasonable but genuine mistake defence.[291] The House of Commons Home Affairs Committee reviewing the Sexual Offences Bill in 2002-03 welcomed proposals for a more objective test considering, *inter alia*, that it was not unreasonable to expect people to take care that another party is consenting to sexual activity, as the cost of so doing is very slight whilst the cost for the victim of failing to do so is severe,[292] and the government took a strong stance on reform generally.[293]

The original draft of the Sexual Offences Bill proposed a two-stage test which considered, first, whether a reasonable person would have doubted the complainant's consent in all the circumstances and, second, whether the defendant acted in a way that would be regarded as sufficient by a reasonable person. However, this approach was considered to be unduly complex and difficult for a jury to operate whilst, further, 'it was feared that it

[290] *Director of Public Prosecutions v Morgan* [1976] AC 182.
[291] See Home Office, *Setting the Boundaries: Reforming the law on sex offences* (Home Office 2000), 23 – 26; David W. Selfe, 'Rape: *Mens rea* and reasonable belief' (2013) 214 *Criminal Lawyer* 3, 3 – 4.
[292] House of Commons Home Affairs Committee, *Sexual Offences Bill* (HC 639, Fifth Report of Session 2002-03), 8.
[293] See further David Ormerod, John Cyril Smith and Brian Hogan, *Smith and Hogan's Criminal Law* (13th ed. Oxford University Press 2011), 743 – 745.

would lead to injustice in some cases because it (arguably) failed to take account of the defendant's particular characteristics, for example a learning disability.'[294] The Chairman of the Criminal Bar Association instead suggested a test which requires the jury to consider 'what a reasonable person "*sharing the characteristics of the defendant*" would have thought',[295] which comes incredibly close the second limb of the hybrid test defended throughout this thesis. The government rejected this, however, on the basis that the jury would be required to take into account *all* of the defendant's characteristics, whereas there a some which should not absolve a person of guilt, such as merely being quick to temper or unable to resist attractive women.

The government considered these concerns and returned with the proposal that now appears in the Sexual Offences Act 2003, namely that the defendant did not reasonably believe in the victim's consent[296] and that whether or not the belief was reasonable ought to be determined having regard to all of the circumstances, including steps taken by the defendant to positively ascertain consent.[297] Crucially, the Minister of State in the Home Office who introduced this amendment commented:

> '[T]he revised version of the reasonableness test moves away from the concept of the "reasonable person" and requires the prosecution to prove that the defendant did not have a reasonable belief in consent. The test is supported by an explanation of the type of criteria to be used to determine whether the defendant's belief in consent was reasonable in relation to the alleged offence. The jury is directed to *have regard to all the circumstances at the time,* including any steps… that the defendant may have taken to establish that the complainant consented to the sexual activity.'[298]

[294] House of Commons Home Affairs Committee (2002-03), 9.
[295] *Ibid*.
[296] Sexual Offences Act 2003, s. 1(1).
[297] *Ibid*., s. 1(2).
[298] House of Commons Home Affairs Committee (2002-03), 10 (per Baroness Scotland of Asthal).

The Home Affairs Committee considering the draft bill approved this amendment, commenting:

> '[T]he revised "reasonableness test" for a defendant's belief in consent is both clearer and simpler than the original drafting. More importantly, *it also addresses the concerns about the potential injustice of applying a "reasonable person" standard to all defendants, regardless of their individual characteristics*. By focussing on the individual defendant's belief, the new test will allow the jury to look at characteristics – such as a learning disability or mental disorder – and take them into account.'[299]

The attention of the Court of Appeal was drawn to these passages in *R v B*; however, the Court considered that the report could not be read as representative of the view of all of Parliament (being derived from a committee thereof) whilst, as a matter of statutory construction, it was not appropriate to take such preparatory material into consideration. As indicated above, whilst the Court cannot be faulted for its technical approach to statutory construction, it is readily arguable that the interpretation of the law arrived at is far harsher than that intended by the parties most intimately involved in its drafting, analysis and debate within Parliament.[300]

With regards to one final criticism, the Court of Appeal maintained that even had the defendant's belief in his partner's consent been directly induced by his illness, for the reasons discussed this could not render an irrational belief as being reasonable. Therefore, unless the defendant's mental illness amounted to the legal defence of insanity, beliefs in consent arising due to mental illness must nonetheless be judged by objective standards of reasonableness which do not take into account the mental illness giving rise to the very beliefs being judged. Firstly, the rule espoused by the Court of Appeal here has no application within the hypothetical facts upon which it is based because, if the defendant falsely believed in his victim's consent because of the effects of his mental illness, this

[299] *Ibid*., 10.
[300] See further Natalie Wortley, 'Reasonable belief in consent under the Sexual Offences Act 2003' (2013) 77(3) *Journal of Criminal Law* 184, 187 – 188.

would be a default of reason arising from a disease of the mind and thus would *prima facie* amount to legal insanity in any event.[301] Secondly, whereas it is accepted that the Court of Appeal is correct to find that the *reasonableness* of a belief is objective, this does not mean that the defendant's mental illness becomes entirely irrelevant to his possession of the belief itself, otherwise the same problems arise as were found with the purely objective *Caldwell* test for recklessness.

*

How, then, does the hybrid approach to reasonable belief remedy this contention, whereby excluding mental illness from consideration as directed in *R v B* risk wrongly convicting defendants with a purely objective test who, because of their individual capacities, would never be able to meet the objective standard of reasonableness? The first limb of the test asks whether the whole circumstances, objectively considered, supports the *reasonable conviction that the defendant's partner was consenting to have sex?* If, for a moment, the term "reasonable" is disregarded – (as if the offence of rape could be committed with the absence of a mere belief alone in the victim's consent) – it might indeed be argued that the circumstances support such a mere conviction. Despite initial protestations, the defendant's partner seemingly undressed herself and offered little resistance to his advances or attempts to escape, and it might be argued that such facts could support somebody's conviction that consent existed.

Of course, this perspective entirely misses the crucial distinction between submission and consent,[302] and it is here that the requirement for any belief to be *objectively* reasonable plays its role. Returning this requirement into consideration, whilst the circumstances might arguably have supported a mere belief in consent, the victim's initial protests, the abusive context of the relationship and the earlier events of forced-feeding clearly undermine any claim that such a conviction in consent could be reasonable. By all accounts, the victim was submitting and not consenting, and this is rape. Thus, the

[301] John J. Child and G. R. Sullivan, 'When does the insanity defence apply? Some recent cases' (2014) 11 *Criminal Law Review* 788, 793.
[302] *R v Olugboja* [1982] QB 320, 332.

reasoning of the Court of Appeal also stands, and the defendant's mental illness has no bearing on the question of whether or not any belief he formed was, in fact, *objectively* reasonable. However, it is submitted that this is not the same as stating that the defendant's mental illness is entirely irrelevant *per se*, unless and until it amounts to the high threshold of insanity. Rather, under the hybrid test, the fact of the defendant's mental illness becomes a relevant consideration for rebutting the second limb.

The second limb of the hybrid test asks *whether it is reasonable to expect anybody in the same circumstance to appreciate the absence of a reasonable conviction in consent?* On the one hand, because the reasonableness of the conviction has already been objectively determined under the first limb of the test, the second limb is presumed to conclude in favour of the prosecution, because it will ordinarily be reasonable to expect anybody – *i.e., the reasonable man* – to appreciate that objectively reasonable conviction. On the other hand, it is under this second limb that the defendant argues that, because of his mental illness and the defects caused to his reasoning or reading of the victim's signals, it would not be reasonable to expect anybody sharing those same characteristics to have realised that the victim was not consenting. Thus, once the standard of reasonableness under the first limb of the test has been proven, the second limb of the test becomes a *rebuttable* presumption.

Indeed, as is demonstrated in section 11.3.4 of this thesis, below, it is under this second limb that the defence of insanity would clearly become relevant, as potentially in the instant case of *R v B*. However, the prosecution would likely triumph under the hybrid test as it did in the original trial and upon appeal. Crucially, the psychiatric evidence affirmed that although the defendant held delusional beliefs regarding his sexual healing powers or connection with God, this did not impact upon his capacity to appreciate yes and no, right and wrong, and the existence or absence of consent. Thus, the second limb of the hybrid test has not been rebutted and remains affirmative such that, concluded together, the circumstances do not support the reasonable conviction that the defendant's partner was consenting, and it is reasonable to expect anybody in the defendant's same circumstances to have appreciated the absence of that reasonable conviction.

It is pertinent to note that the offence of rape and similar sexual offences state the requirement of reasonable belief in the negative, *i.e.,* the offence is committed when the defendant *does not* reasonably believe in consent. As may be discerned from the preceding discussion, this approach only negatives the first limb of the hybrid test such that the prosecution seeks to prove that a reasonable conviction is *not* supported on the facts whilst the defendant is interested in showing the converse. However, the second limb of the hybrid test does not reverse; it is still in the defendant's favour to prove that it is *not* reasonable to expect anybody in the same circumstances to appreciate the lack of reasonable conviction, until which point the second limb is rebuttably presumed in favour of the prosecution. Finally, where the Court of Appeal seemingly left the door open for certain subjective characteristics of the defendant to be taken into consideration vis-à-vis his reasonable belief, it is inescapable that further litigation will be forthcoming on this point as a ruling is required for each and every mental illness that is pleaded.[303] It is submitted, once again, that the hybrid approach to reasonable belief provides a clearer and more rational framework within which a defendant's subjective characteristics may be taken into consideration, whilst preserving the principle that an irrational or delusional belief cannot be reasonable in law.

### *10.3.3. Final Comments on Knowledge, Belief and Suspicion*

It is submitted that knowledge, belief and suspicion describe a spectrum representing the likelihood with which a set of facts or circumstances may objectively be said to exist. Knowledge reflects the truth or certainty of given facts; belief reflects conviction as to the existence of given facts, something less than knowledge yet which remains the only likely or reasonable conclusion on the evidence available; and suspicion reflects the conjecture or mere possibility that given facts may exist. This spectrum is reflected in the first limb of the hybrid tests for knowledge, belief and suspicion, as defined respectively by certainty, conviction and conjecture.

[303] Ronnie D. Mackay, 'R v B: Rape – Consent – Defendant suffering from mental illness at time of offence Court of Appeal' (2014) 4 *Criminal Law Review* 312, 314.

The concept of knowledge has previously been broken down into various sub-categories which, it is further submitted, may each readily be subsumed within the hybrid formulations presented in this section of the thesis. Beginning with the categorisations offered by Duff,[304] explicit knowledge consists of that which is known prominently in the forefront of a person's mind. This concept is reflected simply in the hybrid test – knowledge is itself defined objectively as the certainty that a particular circumstance exists in the first limb of the test, whilst the second limb asks whether or not it is reasonable to expect anybody sharing the defendant's subjective characteristics to appreciate that particular certainty. Tacit knowledge consists of that which is known and readily recallable, but is not necessarily prominent in the forefront of somebody's mind. Little *legal* distinction has been drawn between explicit and tacit knowledge in jurisprudence, each of which may satisfy the orthodox subjective approach to *mens rea* and, as such, each of which equally fall within the same hybrid objective / subjective formulation of the *mens rea* of knowledge.

Latent knowledge consists of those facts which comprise a person's general understanding of the world, facts and / or circumstances, but which does not arise to the level of explicit or even recallable tacit knowledge in the moment of deciding and acting. Whereas the courts have previously expressed scepticism regarding whether or not latent knowledge can form the basis of criminal responsibility such as in *Russell*, more consistent authority across *Bello, Buswell*, *Martindale* and *McCalla* suggests that latent knowledge can indeed suffice. In particular, the courts are reticent to ignore latent knowledge where a defendant claims that they had simply forgotten (and, therefore, no longer *knew*) a particular fact at the time of an alleged offence. A defence of mere forgetfulness would arguably be untenable in law, as it could be raised simply by a defendant in every case requiring the *mens rea* of knowledge, and would often be practically incontestable by the prosecution which would struggle to prove whether or not a person had subjectively remembered a fact. Moreover, a defence of forgetfulness would have the further effect of arbitrarily punishing those defendants with good memory whilst absolving those with poor memory.

[304] Duff (1983).

Following the more consistent authority, latent knowledge is also readily encapsulated within the hybrid objective / subjective approach to *mens rea*. However, in cases such as *Russell*, it would remain open for the jury to consider whether or not it was reasonable to expect anybody to appreciate a certain fact, where particular circumstances might be appreciated for impacting on any reasonable person's faculties of memory. That is to say, whether or not latent knowledge were accepted in any given case would fall within the jury's inherent role of setting the boundaries of reasonable and unreasonable conduct or, in these circumstances, reasonable and unreasonable forgetfulness. Nevertheless, the weight of authority suggests that a defence of mere forgetfulness will typically not be accepted by the courts as negating knowledge that it was otherwise reasonable to expect a person to possess in the circumstances.

Devlin J also offers sub-categories of knowledge in *Roper* which, it is again submitted, can readily be incorporated into the hybrid approach to *mens rea*. First, actual or direct knowledge is effectively equivalent to Duff's explicit knowledge and is thus treated the same within the hybrid formulation of knowledge proposed in this section of the thesis. Second, wilful blindness is also generally accepted in law to fall within the *mens rea* of knowledge, applying the principle that a person should not be permitted to escape the imputation of knowledge when they suspect a given matter to be true and decline to take any readily available steps to confirm or refute that suspicion. The inclusion of wilful blindness within the *mens rea* of knowledge is demonstrated by comparing applications of the hybrid approaches to knowledge on the one hand, and suspicion plus the failure to make readily available inquiries on the other.

This may be exemplified from the facts of *Griffiths* discussed in section 10.3.2.5, above. If the *mens rea* of knowledge is applied, the first limb of the test is satisfied because it was indeed certain that the candlesticks in question were stolen. Regarding the second limb of the test, the circumstances relevant to the defendant included such facts as his own suspicion that the goods were stolen, the manner in which he bought them from a random man on the high street, and the mixed accounts for the provenance of the candlesticks that he gave both to prospective buyers and the police. Thus, it is argued that it would be reasonable to expect any defendant in the same circumstances to appreciate

the fact that the candlesticks were stolen; it would be *unreasonable* for the defendant to claim that he suspected the goods to be stolen but did not wish to ask the man selling them on the high street because he felt inappropriate in so doing or, even worse, did not care to hear the answer. Alternatively, breaking down wilful blindness into its constituent parts, there first existed the conjecture that the goods were stolen, which the defendant himself admitted to suspecting. Second, it was readily open for the defendant to inquire as to the provenance of the candlesticks from the man on the high street, but he did not proceed to do so because he did not consider it to be acceptable. Thus, it is readily appreciable how the concept of wilful blindness falls into the hybrid approach to knowledge, applying the principle that being wilfully blind to certain facts will generally be found to be unreasonable under the second limb of the hybrid test.

The third category of knowledge offered by Devlin J is that of constructive knowledge, which the Divisional Court in *Atwal v Massey* denied had any role to play within the *mens rea* of knowledge itself. That notwithstanding, numerous statutory offences have introduced objective alternatives for knowledge, belief and suspicion, such as "reasonable grounds" for the same, introduced in section 10.3.1 of this thesis, above. The hybrid approach to knowledge, belief and suspicion deals with this by introducing the condition of reasonableness into the first, objective limb of the hybrid test – *i.e.,* the relevant certainty, conviction or conjecture must itself be objectively reasonable in the circumstances. Subsequently, the second, subjective limb of the test becomes a rebuttable presumption in favour of the prosecution as, barring some specific defence, it will be reasonable to expect any normal individual (*i.e.,* the objective "reasonable man") to appreciate a certainty, conviction or conjecture that is itself objectively reasonable. Thus, just as objective "reasonable grounds" for belief or suspicion is generally easier to satisfy than the purely subjective belief or suspicion *simpliciter* under the current law, so the hybrid formulation of "reasonable grounds" which places a reasonableness standard within the first limb of the test will generally be easier to satisfy than the equivalent hybrid *simpliciter* formulation which considers reasonableness under the second limb of the test.

Finally, a number of offences include a negative version of knowledge, belief or suspicion, such as the offence of rape which includes the *absence* of reasonable belief in the victim's

consent. Considering that, in such cases, a standard of reasonableness has been introduced into the first limb of the test, the practical effect is that a defendant will be interested in proving that a belief (conviction) in the victim's consent was objectively reasonable, whereas the prosecution will be interested in proving that that belief was not objectively reasonable. Again, as "reasonableness" has now been considered under the first limb of the test, the second limb is rebuttably presumed in favour of the prosecution, whilst the defendant will seek to rebut that presumption and show that it was not reasonable to hold that expectation of anybody sharing their subjective circumstances. Thus, it is submitted that the hybrid approach to knowledge, belief and suspicion is capable of encapsulating each of actual, tacit and latent knowledge, wilful blindness, reasonable grounds for belief and suspicion, and the absence of such reasonable grounds, all within the objective / subjective formulations of each form of *mens rea*.

## *10.4. Dishonesty*

Under the revised hybrid formulation, dishonesty is defined objectively as occurring when *the defendant's conduct was dishonest by the standards of ordinary people*, and is assessed with the defendant's relevant subjective circumstances in mind by asking, *is it reasonable to expect anybody in the defendant's circumstances to appreciate that dishonesty*?

### *10.4.1. Dishonesty – Subjectivity and Objectivity (Again)*

Dishonesty is a relatively modern concept in English law introduced by the Theft Act 1968; prior to this legislation, "fraudulence" was an approximately equivalent requirement under the Larceny Act 1916. At common law, the mental component of larceny is historically described as being *animo furandi* or "felonious intent." Blackstone's seminal 18th Century treatise, *Commentaries on the Laws of England*,[305] states:

[305] William Blackstone, *Commentaries on the Laws of England in Four Books, Volume 2* (George Sharswood (ed.), J. B. Lippincott Co. 1875).

> 'This requisite, besides excusing those who labour under incapacities of mind or will... indemnifies also mere trespassers and other petty offenders. As, if a servant takes his master's horse without his knowledge and brings him home again; if a neighbour takes another's plough that is left in the field and uses it upon his own land and then returns it; if, under colour of arrear of rent where none is due, I distrain another's cattle or seize them; all these are misdemeanours and trespasses, but no felonies. The ordinary discovery of a felonious intent is where the party doth it clandestinely, or, being charged with the fact, denies it. But this is by no means the only criterion of criminality; for in cases that may amount to larceny the variety of circumstances is so great and the complications thereof so mingled that it is impossible to recount all those which may evidence a felonious intent or *animum furandi*; wherefore they must be left to the due and attentive consideration of the court and jury.'[306]

As Steel identifies, this conception of felonious intent is presented as an 'undifferentiated compound element' which comprises of various factors considered relevant to theft and other property offences, including 'requirements of legal capacity, intent to permanently deprive and lack of a claim of right.'[307]

Dishonesty is not defined explicitly under the Theft Act 1968; however, the act provides a number of negative examples of conduct that is not considered to be dishonest, some of which are clearly traceable to Blackstone's compound account of *animo furandi*. For example, the relevance of a claim of right over property that is alleged to be stolen is reflected in sections 2(1)(a) and (b) of the 1968 Act, which provide respectively that it is not dishonest for a person to appropriate property over which they believe they have a legal claim of right, or where they believe that the property owner would consent. Under both circumstances, the defendant is operating under a belief regarding their legal rights that has subsequently not been considered to reflect either dishonesty or its precursor felonious intent. Indeed, by the late 19th century, many jurists – including members of a Royal Commission considering codification of the criminal law – regarded that felonious

[306] *Ibid.,* 232.
[307] Alex Steel, 'The meanings of dishonesty' (2009) 38(2) *Common Law World Review* 103, 104.

or fraudulent intention effectively consisted of the absence of a claim of right,[308] as similarly defined in sections 2(1)(a) and (b) of the Theft Act 1968.

For much of the life of the *mens rea* of dishonesty, the principal description has been provided by the case of *R v Ghosh*[309] which defined dishonesty with two prongs. First, the jury must decide 'whether according to the ordinary standards of reasonable and honest people what was done was dishonest.'[310] If the defendant's conduct was so dishonest according to this objective limb of the test, the jury decide second 'whether the defendant himself must have realised that what he was doing was by those standards dishonest.'[311] This combination of an objective test alongside a subjective test was intended to avoid the excesses of relying upon either test alone. The Court of Appeal used the hypothetical example of a foreigner who failed to pay the fare for using a bus after arriving from a country where public transport is free. The Court argued, on the one hand, that a purely objective test might unfairly punish such a person for failing to pay for the bus, which would generally be regarded as dishonest by the standards of ordinary people. On the other hand, a purely subjective test would be more difficult to satisfy the more peculiar or disturbed an individual's moral compass, permitting any behaviour that they did not themselves realise would be considered to be dishonest by others.[312]

The parallels between the *Ghosh* test and the hybrid objective / subjective test proposed in this thesis should be plain; the first limb is all but identical, asking whether the defendant's conduct is objectively regarded as dishonest. The second limb marks a crucial variation from the *Ghosh* test, however; the second limb is entirely subjective under the original test, asking whether or not the defendant actually realised that their conduct was dishonest by the standards of ordinary people. However, this invites the problems of subjectivity that were discussed at length in section 8.2 of this thesis, above. The revised second limb of the test avoids these challenges by asking, not whether a particular defendant *actually* appreciated the objective dishonesty of their actions but, whether or not anybody in the same circumstances is *reasonably expected to appreciate* the

[308] See James Fitzwilliam Stephen, *A History of the Criminal Law of England – Vol III* (Macmillan 1883), 131; James Fitzwilliam Stephen, *A General View of the Criminal Law of England* (2nd ed. Macmillan 1890), 146.
[309] *R v Ghosh* [1982] QB 1053.
[310] *Ibid*., 1064.
[311] *Ibid*.
[312] Andrew K. W. Halpin, 'The test for dishonesty' (1996) (May) *Criminal Law Review* 283, 286.

dishonesty of their actions. This revised test continues to capture crucial circumstances and characteristics subjective to the defendant, whilst retaining at its core an objective question of what is reasonably expected from members of society, thus avoiding the problems of imposing a purely subjective test under the second limb of *Ghosh* dishonesty.

The *Ghosh* test stood resilient for 35 years despite attracting some notable criticism, for example, for being unduly complex, for the second limb of the test being partially redundant, or even for the two limbs being self-defeating.[313] Some of the key criticisms fell to be considered by the Supreme Court in *Ivey v Genting Casinos (UK) Ltd. t/a Crockfords*,[314] a civil case concerning the inclusion or otherwise of dishonesty within the meaning of "cheating" in the context of gambling contracts, where the Court nonetheless took the opportunity to redefine dishonesty for the criminal law also.[315] The principal objection considered by the Court followed the argument that 'the less the defendant's standards conform to what society in general expects, the less likely he is to be held criminally responsible for his behaviour.'[316] The Court was further minded that the civil law had settled upon an objective approach to dishonesty with no logical or principled reason why the meaning of dishonesty ought to differ between civil and criminal cases.[317]

The Supreme Court in *Ivey* therefore adopted an objective test for dishonesty, defined simply according to the objective standards of ordinary people. However, reminiscent of the Supreme Court's *dicta* concerning reasonable cause for suspicion in *Lane and Letts*,[318] the Court asserted that this test is not, in fact, entirely objective.[319] Returning to the

[313] For example, see Glanville Williams, 'The standard of honesty' (1983) 133 *New Law Journal* 636; Kenneth Campbell, 'The test of dishonesty in *R v Ghosh*' (1984) 43(2) *Cambridge Law Journal* 349; Edward J. Griew, 'Dishonesty: The objections to *Feely* and *Ghosh*' (1985) *Criminal Law Review* 341.
[314] *Ivey v Genting Casinos (UK) Ltd. t/a Crockfords* [2017] UKSC 67.
[315] See generally, Zach Leggett, 'The new test for dishonesty in criminal law – Lessons from the Courts of Equity?' (2020) 84(1) *Journal of Criminal Law* 37.
[316] *Ivey* [2017], [58].
[317] *Ibid*., [62] – [63].
[318] *Lane and Letts* [2018], [24].
[319] See further Mark Thomas and Samantha Pegg, 'Clarifying the applicable test for dishonesty and modifying *stare decisis*, but otherwise a missed opportunity' (2020) 84(4) *Journal of Criminal Law* 385, 387 – 389; David Ormerod and Karl Laird, 'The future of dishonesty – Some practical considerations' (2020) 6 *Archbold Review* 8.

hypothetical example of the foreigner who failed to pay his bus fare because he mistakenly believed public transport to be free (as in his home country):

> '[T]he man in this example would inevitably escape conviction by the application of the (objective) first leg of the *Ghosh* test. This is because, in order to determine the honesty or otherwise of a person's conduct, one must ask what he knew or believed about the facts affecting the area of activity in which he was engaging. In order to decide whether this visitor was dishonest by the standards of ordinary people, it would be necessary to establish his own actual state of knowledge of how public transport works. Because he genuinely believes that public transport is free, there is nothing objectively dishonest about his not paying on the bus… "dishonestly", where it appears, is indeed intended to characterise what the defendant did, but in characterising it one must first ascertain his actual state of mind as to the facts in which he did it.'[320]

Once again, parallels between the revised objective approach to dishonesty in *Ivey* and the hybrid objective / subjective approach defended in this thesis may be appreciated; under the hybrid test, the jury must similarly take into account relevant subjective characteristics and circumstances of the defendant in considering the second limb of the test. However, the purpose of this subjective inquiry in *Ivey* is to establish, *inter alia*, the defendant's state of mind which is in turn adjudicated objectively according to the standards of ordinary and reasonable people. Under the proposed hybrid test, the first limb remains entirely objective, asking whether the defendant's conduct is dishonest by the standards of ordinary men. The defendant's subjective characteristics and circumstances become relevant under the second limb of the hybrid test in so far as they relate to anybody's capacity to appreciate the objective dishonesty of their actions. It is crucial to reiterate, however, that this is not simply a reproduction of the *Ghosh* test; the second limb of the hybrid test does not require proof that a defendant *actually* appreciated that their conduct would be considered dishonest by the ordinary standards of reasonable

[320] *Ivey* [2017], [60]; approving *Royal Brunei Airlines Sdn Bhd v Tan* [1996] 2 AC 378.

men, but that they had the requisite *capacity* to make this realisation and, in all the circumstances, it was reasonable to expect that any defendant would exercise that capacity.

Whilst regarded by many as a welcome improvement upon *Ghosh*, the revised objective test for dishonesty in *Ivey* has not escaped criticism itself.[321] Clough[322] and Galli[323] each submit that the *Ivey* test successfully deals with the most common criticism of the *Ghosh* test, namely the perverse situation whereby a defendant with a more warped or disturbed sense of dishonesty became more likely to evade the reach of the test – *i.e.,* the "Robin Hood" defence. The hybrid test supported in this thesis similarly addresses this problem; even if an individual defendant has a sense of morality that is severely perverted in contrast to the standards of ordinary people, this does not render it any less reasonable to expect that anybody in the same circumstances would appreciate that their own standards of dishonesty have deviated from those of the society around them, even if convinced that they were nonetheless in the right.

The *Ivey* test has been duly criticised for introducing a greater degree of uncertainty into the law because the 'breadth of the concept [of dishonesty] increases the risk of different courts reaching different verdicts on essentially similar facts, and leaves room for the infiltration of irrelevant factors' when considering whether the defendant's conduct (including their personal characteristics, subjective knowledge and beliefs) is dishonest by ordinary standards.[324] In a similar vein, the move to a purely objective test has been criticised for significantly expanding the reach of offences such as theft and conspiracy to defraud, reducing the scope for legitimate defences in offences that are already characterised by their *actus reus* 'to vanishing point which means that they lack… "manifest criminality".'[325]

---

[321] See broadly, David Ormerod and Karl Laird, '*Ivey v Genting Casinos* – Much ado about nothing?' in Clarry D. (ed.), *The UK Supreme Court Yearbook: Volume 9* (Appellate Press 2019).
[322] Joanne Clough, 'Giving up the *Ghosh*: *Ivey (Appellant) v Genting Casinos (UK) Ltd trading as Crockfords (Respondent)*' (2018) 236 *Criminal Lawyer* 2, 3.
[323] Mark Galli, 'Oh my *Ghosh*: Supreme Court redefines test for dishonesty in *Ivey v Genting Casinos*' (2018) 29(2) *Entertainment Law Review* 55, 57.
[324] Jeremy Horder, *Ashworth's Principles of Criminal Law* (9th ed. Oxford University Press 2019), 404.
[325] Ormerod and Laird (2019), 393.

Conversely, the proposed hybrid approach both reintroduces a greater degree of subjectivity into the test for dishonesty whilst still imposing greater limits upon those characteristics and circumstances that are considered under the second limb of the hybrid test than existed under the purely subjective second limb of the *Ghosh* test. Thus, subjective circumstances and characteristics remain relevant in so far as they must relate to the reasonableness of expecting anybody in the defendant's circumstances to appreciate the dishonesty of their actions; that is, their capacity to appreciate the relation of their actions to the particular *mens rea* of the offence for which they are charged, and the presumed capacities for responsiveness to reasons and ordinary self-control. This, it is submitted, is how the hybrid approach reaches a balance between objectivity and subjectivity within *mens rea* generally.

The aforementioned uncertainty has also given rise to further criticism that the *Ivey* test breaches the requirements of fair notice under Article 7 of the European Convention on Human Rights ('ECHR').[326] Article 7 requires that both criminal offences and their requisite components are sufficiently certain 'to enable the citizen to foresee, if need be with appropriate advice, the consequences which a given course of conduct may entail.'[327] Sullivan and Simester,[328] and Dyson and Jarvis[329] each suggest that the *Ghosh* test may have previously achieved this necessary certainty through the second subjective limb of the test, following which a defendant 'could not be dishonest without realising that fact.'[330] This point is far from trite as, in relation to the *mens rea* of dishonesty generally, 'its sphere of operation is enormous' being required for around one-half of all indictable charges tried before the criminal courts,[331] whilst the offences of theft and conspiracy to defraud 'could not be wider in terms of their conduct elements.'[332] Moreover, each of the

[326] *Ibid.*, 394 – 398.
[327] *R v Rimmington* [2005] UKHL 63, [35].
[328] G. R. Sullivan and Andrew P. Simester, 'Judging dishonesty' (2020) 136(Oct) *Law Quarterly Review* 523, 526.
[329] Matthew Dyson and Paul Jarvis, 'Poison Ivey or herbal tea leaf?' (2018) 134(Apr) *Law Quarterly Review* 198, 202.
[330] *Ibid*; citing *R v Pattni* (unreported) [2001] *Criminal Law Review* 570; see also Karl Laird, 'Dishonesty: *R v Barton*; *R v Booth;* Court of Appeal: 29 April 2020' (2020) 11 *Criminal Law Review* 1065, 1068 – 1069; Graham Virgo, 'Cheating and dishonesty' (2018) 77(1) *Cambridge Law Journal* 18.
[331] Horder (2019), 402.
[332] Sullivan and Simester (2020), 526; citing *R v Hinks* [2001] 2 AC 241; *Scott v Metropolitan Police Commissioner* [1975] AC 819.

Law Commission,[333] the Joint Parliamentary Committee on Human Rights,[334] and the Attorney General[335] explicitly envisaged the retention of the objective-subjective *Ghosh* test in the drafting of the Fraud Act 2006, reflected in the explanatory notes to that Act.[336]

For reasons already largely rehearsed, it is submitted that the proposed hybrid approach to dishonesty avoids this criticism of the *Ivey* test concerning certainty and the implications for Article 7 of the ECHR. Although the second limb of the hybrid test does not require that the defendant *actually* appreciated the objective dishonesty of their actions, it does require that the circumstances are such that it would be *reasonable to expect anybody* to make that same realisation – *i.e.,* that they have the capacity to appreciate the nature of their actions as they relate to the *mens rea* of the offence. This means that any characteristics or circumstances subjective to the defendant and which would impact upon the aforementioned capacity are taken into consideration under the second limb of the test, thus having due regard to relevant and potentially exculpatory subjective factors, as with the second limb of the previous *Ghosh* test. More generally, the hybrid approach redresses some of the key criticisms against both the *Ghosh* and *Ivey* tests whilst retaining the benefits of both an objective and subjective approach.

### 10.4.2. ***Testing Hybrid Dishonesty in Jurisprudence***

#### *10.4.2.1. R v Gilks*

The defendant in *R v Gilks*[337] placed bets on horse races at a betting shop; on one occasion, the shop manager overpaid the defendant having mistaken which horse he had bet on. It was accepted by the defendant that he knew he was not entitled to the winnings when he received them, however he contended that he had not been dishonest in keeping that overpayment. Specifically, the defendant submitted that although it would be dishonest to keep too much change given over by a grocer, 'bookmakers and punters are a race apart

[333] Law Commission, *Fraud: Report on a Reference under Section 3(1)(e) of the Law Commissions Act 1965* (Law Com No. 276, 2002), [5.6] – [5.19].
[334] Joint Parliamentary Committee on Human Rights, *Legislative Scrutiny: Sixth Progress Report* (HL 134, HC 955, Fourteenth Report of Session 2005-06), [2.15] – [2.25].
[335] See Sullivan and Simester (2020), 526.
[336] Explanatory Notes to the Fraud Act 2006.
[337] *R v Gilks* (1972) 56 Cr App R 734.

and… when you are dealing with your bookmaker different rules apply.'[338] The Court of Appeal approved of the direction to the jury provided by the trial judge, who instructed that they should 'try to place yourselves in that man's position at that time and answer the question whether in your view he thought he was acting honestly or dishonestly', thereby approving an entirely subjective test.[339] The defendant's appeal therefore failed and his original conviction for theft was upheld, the jury having rejected the defendant's assertion that he subjectively believed himself to have been acting honestly.

The hybrid test would have little difficulty with similarly upholding conviction in the circumstances presented. Concerning the first limb of the test, it is readily arguable that keeping an overpayment which one knows to be such is to be regarded as dishonest by ordinary standards. Certainly, the fact that an overpayment is received from a bookmaker as opposed to a grocer should have no bearing on the honesty or dishonesty of keeping that overpayment. Put differently, it is difficult to argue that it is inherently honest conduct to keep money that others have paid over by mistake and to which the recipient has no legal right of claim. Regarding the second limb of the test, perhaps the only relevant circumstance to consider from the defendant is his subjective belief that his behaviour had not been dishonest, albeit the jury clearly ultimately concluded that this belief had not been genuinely held in the event.

Even had the defendant's account been credible, a *personal* belief in the honesty of keeping an overpayment received from bookmakers does not render it any less reasonable to expect that anybody in the same circumstances would appreciate that such conduct would be considered dishonest *by the standards of ordinary reasonable men*. Indeed, the defendant indicates some recognition of this point when he accepts that retaining an overpayment from a grocer would indeed be dishonest. He therefore clearly recognises something potentially dishonest in the act of retaining an overpayment, even if contending that the identity of the giver of that overpayment is relevant to the honesty of its retention.

[338] *Ibid.*, 738.
[339] *Ibid.*, 738 – 739.

#### *10.4.2.2. R v Feely*

Whereas the Court of Appeal in *Gilks* approved what appears to be an entirely subjective test, the Court of Appeal in *R v Feely*[340] appeared to go significantly in the other direction towards a mostly, if not entirely, objective approach. The defendant in *Feely* was employed as a branch manager within a chain of bookmakers where, in mid-September 1971, the employers issued a circular stating that the practice of borrowing money from the shop tills was thereafter prohibited. In early October, the defendant took around £30 from the shop till to give to his father who was out of work and, at the time, neither informed his employer nor placed any note in the safe to record the withdrawal. A few days later the defendant was transferred to another branch and his replacement noted the shortfall in the safe. At his time, the defendant gave an account of needing to borrow the money and also presented a written note recording that he owed the money to his employer. During his interview with the police, the defendant asserted that he had borrowed the money with the intention of repaying it and, furthermore, that his employer owed him around £70 from which he wanted them to deduct the money that he had borrowed. The defendant was convicted of theft and appealed, *inter alia*, on the grounds that the judge had misdirected the jury with regards to dishonesty.

In the defendant's favour, evidence had been adduced that the borrowing of money from the shop till was a commonplace and accepted practice within his employer's business. Specifically, the defendant submitted that branch managers such as himself were personally responsible for cash deficiencies as a matter of practice, were similarly responsible for advances made to clients, and also for advancing loans or wages to employees, that he provided an "IOU" for any deficiencies to incoming branch managers, and had credit with his employer exceeding the amount that he had borrowed. Against the defendant's favour, however, included that fact that evidence given by his father refuted the reasons for which the defendant had given money to him and which the defendant ultimately accepted had been an untrue story. Equally, it did not favour the defendant that he had failed to inform his employer or leave any note regarding the missing money until he was questioned regarding the shortfall; nor, of course, that his employer had officially prohibited any such practice of borrowing from the shop funds

[340] *R v Feely* [1973] QB 530.

only weeks before the defendant's alleged theft. The Court of Appeal was clearly minded of the defendant's guilt, therefore;

> 'The honest employee who has to deal with an emergency for which cash is necessary there and then usually tells his employers either at the time or shortly afterwards what he has done, whereas the rogue says nothing until his taking is found out whereupon he asserts his intention to repay and stresses his ability to do so.'[341]

The defendant's appeal was successful, however, due to misdirections given by the trial judge. In particular, the trial judge had declined to leave the defendant's defence to the jury – *i.e.*, that he borrowed the money in accordance with common practice and intended to repay the money. The Court of Appeal considered it 'possible to imagine a case of taking by an employee in breach of instructions to which no one would, or could reasonably, attach moral obloquy', providing the example of an employee who borrows 40p of small change in order to pay for a taxi fare because the driver cannot give change for her £5 note, which she then uses to immediately repay the borrowed money.[342]

On the matter of dishonesty specifically, the Court considered the matter to be a question of fact to be determined by the jury applying the standards of ordinary decent people,[343] thus approving a predominantly objective test, albeit one applied to the defendant's subjective state of mind. Whilst the Court asserts that most examples of people taking money from tills or safes to which they have no lawful claim will usually amount to theft, this did not eradicate in law the potential availability of the defence forwarded by the defendant in *Feely* (even if that defendant was found not to be credible and his defence rejected on the available evidence). Three factors which appear to be important for such a defence from the Court's judgment and consideration of other jurisprudence include the relatively low value of money borrowed, the ability for the borrower to repay that money immediately from their own funds as opposed to their expectance of being able to in the

[341] *Ibid*., 536.
[342] *Ibid*., 539; approving *R v Williams* [1953] 1 All ER 1068, 1070.
[343] *Ibid*., 537 – 538.

future, and notification to the party from whom money is borrowed either immediately or shortly thereafter.

In agreement with the Court of Appeal's own consideration of the defendant's guilt (notwithstanding allowing the appeal on technical grounds), it is equally likely that the hybrid approach to dishonesty would support the initial conviction in *Feely*. Beginning with the first limb of the test, it may readily be stated that taking money from another without their knowledge, permission or any other claim of right is generally regarded as dishonest according to the standards of ordinary people. Considering the second limb of the test, it might have been relevant that the defendant asserted his intention to repay the money had his evidence not ultimately lost credibility at trial. So far as his asserted defence is concerned and the factors relevant for consideration, the sum borrowed cannot be considered as insignificant (amounting to more than £400 today) whilst, by his own evidence, the defendant had borrowed the money to cover a shortfall and did not possess his own funds to repay it immediately. Equally, it might have been relevant that there was a common and accepted practice of borrowing money within the company for which the branch managers were personally responsible; however, the defendant's employer had clearly and unequivocally declared this practice to be prohibited only weeks earlier. Therefore, there are no particular circumstances presented in the defendant's case which diminish the reasonable expectation that anybody in the same circumstances would appreciate the dishonesty of their conduct by ordinary standards.

Even had the defendant's credibility not been called into question and his intention to repay the money accepted, it remains likely that he would have been convicted under the hybrid approach to dishonesty. Particularly damning is the fact that the defendant's employer had explicitly prohibited the further borrowing of money by employees, regardless of whatever practice had previously been accepted. Without this, the defendant might fairly have argued that it would be unreasonable to expect anybody in the same circumstances to appreciate the dishonesty of an act of borrowing that was widely and explicitly accepted by his employer. Against such a clear edict from his employer, however, it is difficult for the defendant to rely upon that previous practice of borrowing money in order to claim that it would not be reasonable to expect him to appreciate that

such borrowing was dishonest. The defendant was clearly on notice that such practices were no longer acceptable and, thus, it is reasonable to expect anybody in the same circumstances to appreciate the dishonesty in continuing such a practice without their employer's knowledge and contrary to their express prohibition.

#### 10.4.2.3. *Boggeln v Williams*

*Boggeln v Williams*[344] is often cited as the quintessential example of a case which sits on the borderline between honest and dishonest conduct. The defendant's electricity supply was cut off after he failed to pay his electricity bill, and he reconnected the supply without the requisite authority of the Electricity Board. However, he did inform the Board that he had so reconnected the electricity supply and further ensured that his consumption of electricity was recorded through the meter (which he could have bypassed had he so wished). Furthermore, it was the defendant's assertion that, at the time of reconnecting his electricity supply, he believed that he *would* be able to pay for the power that he was consuming (albeit it is not specified whether or not he *could* have so paid immediately at the time). The defendant was charged and convicted of dishonestly using electricity without lawful authority,[345] but his conviction was overturned upon first appeal on the basis that his belief in his ability to pay had not been proven to be unreasonable and, therefore, that his state of mind at the time had not been dishonest. The prosecutor further appealed on the ground that a person's belief concerning their own honesty or dishonesty was an irrelevant consideration.

An important finding of fact at trial was that the defendant did not believe that the Electricity Board had consented to his reconnecting the supply, but he nevertheless did believe that he was not acting dishonestly on account of 'giving notice of his intention and by ensuring that consumption was duly recorded through the meter.'[346] Moreover, it was found that the defendant could have bypassed the meter if he so wanted and, furthermore, it was accepted at trial that the defendant had genuinely intended to pay the electricity that he consumed. It is further relevant to note section 2(2) of the Theft Act

[344] *Boggeln v Williams* (1978) 67 Cr App R 50.
[345] Theft Act 1968, s. 13.
[346] *Boggeln* (1978), 53.

1968 which provides that an appropriation of property belonging to another may still be regarded as dishonest notwithstanding that the person is willing to pay for that property.

The Court of Appeal regarded that the offence for which the defendant was convicted consisted of using electricity *both* dishonestly *and* without authority. Consequently, the Court determined that the 'fact that the defendant did not believe at the time he re-connected his supply that he had the consent of the Board does not of itself make the defendant's conduct dishonest in law', this being a question of fact to be determined by the jury. Furthermore, the Court appeared to distance itself from *Feely* in supporting a subjective assessment of dishonesty relating to the defendant's actual state of mind as opposed to the standards of ordinary people. The prosecutor's appeal was therefore dismissed and the defendant's appeal and subsequent acquittal was maintained.

The facts of *Boggeln* arguably pose the greatest challenge to the first limb of the hybrid test – that is to say, is the action of reconnecting an electricity supply that has been disconnected for want of payment, but whilst informing the Electricity Board of the intention to so reconnect that supply, and ensuring that that supply is correctly metered to facilitate future payment, an objectively dishonest action? To attempt to simplify this, the fact that the defendant's supply had been disconnected for want of payment may be withdrawn as, following the Court of Appeal's determination, the absence of due authority to use electricity is separate from the question of dishonesty and was conceded by the defendant in any event. The facts that the defendant both informed the Electricity Board and ensured that the electricity supply was metered – things that he actually *did* as opposed to matters subjectively in his mind – remain relevant in considering whether his actions overall meet an objective definition of dishonesty by ordinary standards.

Whilst *Boggeln* is indeed a difficult borderline case to so determine, it is suggested that the defendant's conduct would not be considered dishonest under the first limb of the hybrid test. Certainly, his actions would have been dishonest if done in secrecy, but the defendant had been open and frank with the Electricity Board; equally, his actions would have been dishonest if he had bypassed the electricity meter, but he had not done this either. Indeed, from the facts presented, each of the defendant's actions are commensurate

with an individual intending to pay for electricity, ensuring the means of recording his usage, and placing his provider on notice to bill for that usage, notwithstanding that he lacked actual authority to reconnect the electricity supply. Put differently, the facts do not at all suggest that the defendant was attempting to obtain any undue advantage from the Electricity Board or obtain any supply of electricity without due payment; the defendant's conduct did not attempt to deceive the Electricity Board, nor did he act covertly or secretively. It may therefore be concluded that the defendant's conduct does not meet an objective description of dishonesty according the standards of ordinary people, and his acquittal is consequently supported.

Supposing the afore conclusion to be incorrect and the first limb of the hybrid test is satisfied in *Boggeln*, the second limb of the hybrid test becomes harder still to determine. Phrased to include the pertinent subjective circumstances the question follows, if a person reconnects their disconnected electricity supply with the genuine intention of paying for that supply and belief in their ability to do so, whilst further informing his supplier that he has so reconnected his supply, and ensuring that that supply is correctly metered, is it reasonable to expect that person to appreciate that his conduct has been dishonest by the standards of ordinary people? On the one hand, there are no circumstances subjective to the defendant in the present case to suggest that the he lacked any *capacity* to appreciate the dishonesty of his actions by the standards of others, even if he mistakenly believed in his own honesty. Following this argument, the fact that the defendant may have genuinely intended to pay for his electricity supply and was honest in informing the Board of its reconnection does not necessarily mean that it is unreasonable to expect him to appreciate that his actions are nonetheless dishonest by the standards of ordinary people – his capacity for this understanding remains unaffected.

On the other hand, however, it might be argued that the fact that the defendant informed the Electricity Board that he was reconnecting his supply may be determinative of the second limb of the hybrid test. Specifically, if a person gives a full, honest and open account of their intended actions and then proceeds to act in complete accordance with that account, it is difficult to maintain that they have been in any way deceitful, untrustworthy, insincere or misleading regarding that conduct. It may, therefore, arguably

be unreasonable to expect somebody who has been honest and sincere regarding their conduct to appreciate that they are nonetheless regarded as dishonest by the standards of others.

Charged with acting dishonestly in such circumstances, anybody could (and likely would) respond, "At which point was I dishonest, when I stated honestly to the affected party what I intended to do and then so acted in accordance with that honest statement?" The argument follows that it is not reasonable to expect anybody to appreciate the objective dishonesty of their actions when they have acted in accordance with a prior statement of intent. The crucial and exculpatory point under this argument is not that the defendant intended to pay for the electricity supply, nor that he subjectively believed his own conduct to be honest; rather, it is in the fact that he took extensive measures to ensure that his conduct would comply with ordinary standards of honesty, not least by providing prior notice of his intended actions and, by virtue of his subsequent actions conforming with that notice, that notice was itself honest.

The difficulty with this latter argument is that it is arguably circular and returns the question to the first limb of the test. That is to say, in arguing that the second limb of test is not satisfied because the defendant has been honest in giving prior notice of his intended actions and has acted accordingly, the argument in fact being made is that the defendant's conduct was not objectively dishonest to begin with because of the fact that he provided the prior honest notice of intent. If the first limb of the hybrid test has already been satisfied, however, this latter circular argument must fail and the defendant in *Boggeln* would be convicted under the hybrid test. That notwithstanding, it is reiterated that the first limb of the hybrid test likely would not be satisfied in the present case, thus leading to acquittal in concurrence with the defendant's two successful appeals. Furthermore, it is reiterated that *Boggeln* provides the quintessential borderline case and the difficulty with which the hybrid approach to dishonesty deals with the case is similarly experienced applying both the *Ghosh* and *Ivey* tests also, such difficulty arising in particular under the objective limb of each of the tests discussed.

#### *10.4.2.4. R v Ghosh*

As discussed in the introduction to this section of the thesis, the case of *R v Ghosh*[347] provided the longstanding test for dishonesty for more than three decades which, providing inspiration for the hybrid test adopted in this thesis, asks first whether the defendant's conduct was dishonest according to the standards of ordinary men and, second, whether the defendant actually appreciated the fact that his conduct was dishonest by those standards. The defendant was a surgeon who, whilst acting as a *locum tenens* consultant at a hospital, claimed fees for a number of operations which had either been performed by a different surgeon or under National Health Service provisions, and to which he was consequently not entitled. The defendant was convicted on one count of attempting to procure by deception the execution of a valuable security[348] and three counts of obtaining money by deception.[349] The defendant's case was that there had been no deception because the fees paid to him were either legitimately due for consultations under the relevant regulations or were otherwise the balance of fees that were properly due.

The jury found the defendant guilty on all four counts thereby clearly rejecting both his bare denial of dishonesty and his assertion of having been legitimately entitled to the sums claimed. The defendant appealed against the direction given by the trial judge on the question of dishonesty resulting in the *Ghosh* test that has been discussed earlier in this section of the thesis. Crucially, however, the Court of Appeal was satisfied that the defendant was dishonest on either an objective or subjective test 'once the jury had rejected the defendant's explanation of what happened.'[350] Thus, despite a successful appeal on the technical legal point of the trial judge's direction, the defendant's conviction was upheld nonetheless.

The same conclusion is inevitable upon the application of the hybrid approach to dishonesty defended in this thesis. First, it is objectively dishonest by all accounts for a person to obtain through false representations payment for work that they have not done

[347] *R v Ghosh* [1982] QB 1053.
[348] Theft Act 1968, s. 20(2).
[349] *Ibid*., s. 15(1).
[350] *Ghosh* [1982], 1064 – 1065.

and / or that had been completed by others. Second, a similar conclusion follows that once the jury had rejected the defendant's bare claim of being entitled to the payments he received; in the absence of any further justification or relevant considerations, it is perfectly reasonable to expect anybody in the same circumstances to appreciate the objective dishonesty of making false representations to receive payment for work that they have not completed. The defendant in *Ghosh* would undoubtedly be convicted under the hybrid approach to dishonesty also.

#### *10.4.2.5. R v Hayes*

The case of *R v Hayes*[351] revealed some of the potential deficiencies within the *Ghosh* test in relation to defendants whose own moral codes and, crucially, appreciation of the standards held by ordinary people were distorted from that ordinary standard. The defendant was an employee at a bank who, within the context of his working duties, conspired with others to manipulate submissions of the Japanese Yen LIBOR which would directly benefit his employer bank and, through bonuses for good performance, would indirectly benefit himself. The defendant was convicted on multiple counts of conspiracy to defraud and appealed on the sole ground the he ought to have been permitted to rely on evidence of practices and ethos within the banking industry in the determination of the first objective limb of the *Ghosh* test. In particular, the defendant submitted *inter alia* that the manipulation of the LIBOR rate was a widespread practice within the industry with many banks and traders offering manipulated rates to obtain an advantage; that his activities had been condoned and encouraged by his own employers; that inherent conflicts existed in the operation of the LIBOR mechanism; and that regulatory bodies such as the Bank of England and the Financial Services Authority were equally aware of the flawed governance of LIBOR.[352]

Whilst accepted that the defendant's preferred evidence would likely be relevant to the second subjective limb of the *Ghosh* test, the Court of Appeal declined to make a similar finding in relation to the first objective limb. The Court approved of the direction of the

[351] *R v Hayes* [2015] EWCA Crim 1944.
[352] See further Jonathan Rogers, 'Dishonesty in the first LIBOR trial' (2016) 3 *Archbold Review* 7, 7 – 8.

trial judge that there are 'no different standards which apply to any particular group of society, whether as a result of market ethos or practice',[353] adding further that the only purpose for including such evidence within the first limb of the test could only be so that the 'jury would be asked to set an objective standard for a market or a group of traders (whatever that standard might be) and not the ordinary standards of honest and reasonable people.'[354] The Court added that it could significantly impede on the proper conduct of business if standards of honesty were to be set by a market as opposed to by reference to the standards of ordinary people. Where history has shown that certain markets adopt dishonest patterns of behaviour from time to time, this should not be interpreted as altering the objective standards of honest and reasonable people but, rather, as that market abandoning those ordinary standards of honesty.[355] The defendant's appeal was therefore dismissed and his conviction for conspiracy to defraud upheld.

Once again considering how the hybrid test for dishonesty might have applied in this case, the same finding as the Court of Appeal is inevitable concerning the objective first limb of the test. Thus, it may fairly readily be concluded that falsely manipulating LIBOR estimates is a patently dishonest course of conduct; where such estimates are supposed to be given in good faith and to establish a fair and balanced playing field in the LIBOR markets, it is inescapably dishonest by ordinary standards to manipulate those estimates in order to obtain an unfair advantage over the market. The defendant's preferred evidence is relevant to the subjective second limb of the hybrid test, although likely would not assist the defendant in the present case; there is little evidence offered which would reasonably impact upon anybody's *capacity* to appreciate that their actions were dishonest by the standards of ordinary men.

The defendant might argue that, in light of a widespread industry practice and particular endorsement by their employer, it would not be reasonable to expect anybody in those same circumstances to appreciate that manipulating the LIBOR was dishonest by objective standards. However, the retort follows that the ordinary, reasonable and honest

[353] *Hayes* [2015], [24].
[354] *Ibid*, [29]; see further Nicholas Dent and Áine Kervick, '*Ghosh*: A change in direction?' (2016) 8 *Criminal Law Review* 553, 554 – 556.
[355] *Ibid*., [32].

man on the street does not necessarily lose sight of the fact that particular conduct is dishonest simply because others are engaged in that same conduct or his employer has condoned it. Thus, in the instant case, it remains reasonable to expect that anybody in the same circumstances as the defendant would nonetheless appreciate that their conduct had been dishonest by ordinary standards.

#### *10.4.2.6. Ivey v Genting Casinos (UK) Ltd.*

The case of *Ivey v Genting Casinos (UK) Ltd*.[356] has an intriguing story which begins with a Chinese heiress who, scorned by a Las Vegas Hotel and Casino that took out a warrant for her arrest after she forgot to repay a guarantee for a friend, vowed to "kill this MGM" and became an expert in a form of cheating called edge-sorting. The story ends with a significant change in the UK law concerning the *mens rea* of dishonesty, overriding a test which had previously stood for more than 35 years.[357] With regards to the particular facts of the case, the claimant was a professional poker player who was introduced to and learned edge-sorting from the aforementioned heiress, and the pair then proceeded to play together at various casinos in the US and UK. The claimant won £7.7 million from one casino in London which subsequently refused to pay out, and for which the claimant sued under the Gambling Act 2005.[358] The defendant casino asserted that the claimant had breached an implied contractual term that neither party would cheat and, furthermore, that the claimant was guilty of the criminal offence of cheating[359] and could not recover his winnings under the principle of *ex turpi causa non oritur actio.*

The claimant was unsuccessful in claiming for their winnings at both first instance and upon appeal, the judge at first instance having found that the claimant had indeed cheated in breach of the implied contractual term. Further, whilst *Ivey* was a civil claim and the claimant had not been charged with the criminal offence of cheating, the Supreme Court determined that the offence does not contain the *mens rea* of dishonesty in any event. Whereas this technically precludes the need for any further analysis of *Ivey* under the

[356] *Ivey v Genting Casinos (UK) Ltd.* [2016] EWCA Civ 1093.
[357] See further Brenda Hale, 'Dishonesty' (2019) 48(1-2) *Common Law World Review* 5.
[358] Gambling Act 2005, s. 42.
[359] *Ibid.*, s. 42(3)(a).

hybrid approach to dishonesty, it may nonetheless be illustrative to consider how the hybrid test might have determined the case. With regards to relevant evidence, the claimant was considerably forthright in giving his account of the technique of edge-sorting, which consists of noticing small manufacturing imperfections on the reverse-side of playing cards which give away their identity; he regarded himself to be an "advantage player" and edge-sorting to be a form of legitimate gamesmanship. That being said, the claimant's technique also required that the same deck of cards be recycled throughout the game and for certain cards to be rotated. The claimant had persuaded the casino croupier to indulge these requirements as part of his purported "superstition", which Tomlinson LJ in the Court of Appeal opined to be deception.[360]

Supposing, then, that the offence of cheating did require the *mens rea* of dishonesty – were Mr. Ivey's actions so dishonest under the hybrid test? Starting with the objective limb, it is arguable on the one hand that the technique of edge-sorting alone is not necessarily dishonest or even cheating – it consists of paying special attention to minor defects on the cards and remembering whether those cards were relatively high or low. On the other hand, the claimant's technique also required re-using the same deck of cards and rotating certain cards, the performance of which the claimant had obtained by lying to the croupier about his superstitions. This went beyond merely paying extra observational attention to the game, but involved actively deceiving another in order to change the manner and rules by which the game was ordinarily played vis-à-vis re-using the decks of cards and rotating cards around. With these additional elements of active deception in order to manipulate the performance of the game, it is more readily argued that Mr. Ivey's actions were dishonest by ordinary standards.

Turning to the second limb of the hybrid test, no particular evidence was offered which calls into question the Mr. Ivey's capacity to appreciate the nature of his actions, albeit it was accepted by the judge at first instance that he honestly believed himself to be an advantage player and edge-sorting not to be cheating. That being said, it must have been clear to Mr. Ivey that the casino croupier might be suspicious when asked to recycle decks and rotate particular cards without any explanation, to the extent that he found it necessary

[360] *Ivey* [2016], [112].

to justify these requests according to a feigned superstition. Equally, it must have been clear that an honest explanation – *i.e.,* to assist with his edge-sorting – would unlikely have been accepted by the croupier, again rendering it necessary to fabricate his superstition as an explanation. Moreover, Mr. Ivey is described as a professional poker player and gambler, from which it might even be expected that he would have a more acute understanding of the line between honest and dishonest behaviour in card games. Even if Mr. Ivey genuinely believed that his actions had not been dishonest, the acts of lying in order to manipulate the manner in which the game was played are likely determinative of the fact that those actions would be considered dishonest by others. It is therefore fair to argue that it would be reasonable to expect anybody in the same circumstances to appreciate that their actions were dishonest by the standards of ordinary people.

### *10.4.3. Final Comments on Dishonesty*

As with the *mens rea* of recklessness in particular, the jurisprudential development of dishonesty has been similarly plagued by the tension between objectivity and subjectivity, encountering similar problems. A predominantly objective approach such as currently exists under the *Ivey* test, like *Caldwell* recklessness, risks failing to take relevant subjective factors into consideration. Meanwhile, a predominantly subjective approach, even in combination with an objective test as in *Ghosh*, perversely risks being undermined in situations where defendants have a particularly distorted moral compass.

The hybrid approach to dishonesty defended in this thesis is notably similar to the *Ghosh* test, but proposes to strike the balance between objectivism and subjectivism differently. The first limb of the test follows the objective limb of the *Ghosh* and *Ivey* tests, requiring simply that the defendant's conduct is regarded as dishonest by the standards or ordinary reasonable people. Under the second limb of the test, however, it is no longer necessary to demonstrate that a defendant *actually* appreciated the dishonesty of their actions; the removal of this purely subjective test seeks to avoid the issues with the *Ghosh* test specifically, as well as broader issues associated with purely subjective *mens rea* discussed in section 8.2 of this thesis, above. Under the revised second limb, it must be

demonstrated that it is reasonable to expect anybody in the same circumstances as the defendant to appreciate the dishonesty of their actions according to ordinary standards. This reformulation thus focuses on a defendant's *capacity* to appreciate the nature of their actions, and not whether a particular, subjectively dishonest state of mind can be proven to have existed.

## *10.5. Negligence*

Under the revised hybrid formulation, negligence is defined objectively as occurring when *the defendant's conduct was unreasonable*, and is assessed with the defendant's relevant subjective circumstances in mind by asking, *is it reasonable to expect anybody in the defendant's circumstances to appreciate the unreasonableness of their conduct*?

Offences of negligence do not traditionally require the prosecution to prove that the defendant possessed any particular culpable state of mind. Indeed, it is often commented that the state of mind of a defendant is immaterial for offences of negligence; however, this point is contested on the basis that a large number of legal defences remain available to crimes of negligence whilst, as the following chapter of this thesis demonstrates, those defences can each be related to one or more of the three capacities relevant for *mens rea* under the hybrid formulation. If, therefore, *mens rea* is not irrelevant insofar as the defendant's capacities are concerned and yet it is not required for the prosecution to prove *mens rea*, the hybrid approach deals with this by introducing a rebuttable presumption that the second limb of the hybrid test is satisfied in relation to crimes of negligence.

The logic behind this presumption should be obvious to appreciate in relation to negligence; if the objective first limb of the test is satisfied by requiring that the defendant's conduct breached an objective standard of *reasonableness*, then it follows as a matter of due course that it is *reasonable* to expect anybody (*i.e.,* the hypothetical reasonable man) to appreciate the fact that their conduct was unreasonable by those standards, unless evidence is adduced (*i.e.,* relating to the defendant's subjective circumstances) which would undermine anybody's capacity to so appreciate that nature of their conduct. Phrased differently, it is not unreasonable to expect ordinary reasonable

people to appreciate when conduct is unreasonable, given that unreasonableness is determined objectively by reference to the (hypothetical) ordinary reasonable person. That having been said, by treating the second limb of the hybrid test as a *rebuttable* presumption, the door remains open for defendants to introduce relevant subjective circumstances and characteristics in order to build a legal defence, whilst the prosecution is *prima facie* not required to prove the second limb of the test.

### *10.5.1. The Reasonable Person and Subjectivity*

Orthodoxy provides that neither offences of negligence nor strict liability require any inquiry into the defendant's subjective state of mind in order for the offence to be proven. Stated more precisely, strict liability offences may be committed where it is not necessary to prove *mens rea* in relation to *at least one* element of the *actus reus* of the offence.[361] However, where some element of *actus reus* lacks corresponding *mens rea* for a strict liability offence, it is similarly unnecessary to prove that the defendant's actions have fallen below any given standard; the commission of the *actus reus* alone is sufficient. Offences that can be committed by negligence also do not require proof of *mens rea*; what matters, however, is the defendant's conduct – 'did the defendant behave in a way which was reasonable in the circumstances.'[362] The applicable test for the requisite standard of care in negligence is that of the hypothetical reasonable person which, again, orthodoxy provides to be an entirely objective test; the hypothetical reasonable man is not coloured with characteristics that are subjective to the individual defendant.

The cases discussed in the following section reveal that, although dominant, the orthodox approach to negligence is not necessarily applied all the time in practice, with examples of age and inferior training being considered in the application of the reasonable man test. Further, there is strong academic support in favour of bringing at least some subjective characteristics into consideration when applying the objective reasonable person test, such as a defendant's age, maturity and intellect, certain mental illnesses or cognitive deficiencies, or inferior training and experience in circumstances requiring particular

[361] John Child and David Ormerod, *Smith, Hogan, and Ormerod's Essentials of Criminal Law* (3rd ed. Oxford University Press 2019), 88.
[362] Jonathan Herring, *Criminal Law: Text, Cases, and Materials* (9th ed. Oxford University Press 2020), 151.

skills. Simester, Spencer, Stark, Sullivan and Virgo[363] offer the hypothetical example of a child who falls into a swimming pool whilst their parent is not attending, suggesting that we might readily hold that parent as negligent for failing to give the same care and attention as a reasonable parent would. Suppose, however, that parent is blind and did not see their child wandering near the pool, or perhaps did not even realise that there was a pool nearby. In such circumstances, they argue, the defendant would not be expected to 'behave as if she were sighted, but rather as a reasonable blind person would.'[364]

A number of further prominent writers, including Ashworth,[365] Horder,[366] Norrie[367] and Hörnle[368] each advocate in favour of greater consideration for relevant subjective factors in the application of an otherwise objective reasonable man test. As Horder summarises, a person who is negligent may be regarded as culpable because they have not taken the necessary care and attention that is required and expected of a reasonable person so as to avoid causing harm; '*so long as the individual had the capacity to behave otherwise at the time,* it is fair to impose liability in those situations where there are sufficient signals to alert the reasonable citizen to the need to take care.'[369]

### *10.5.2. Testing Hybrid Negligence*

Given that offences of negligence do not traditionally require any inquiry into the defendant's state of mind or raise issues relating to *mens rea* and, with some exceptions, are largely concerned with regulatory and other less serious offences, there is relatively little contentious jurisprudence that has developed around these concepts in criminal law that may assist or inform the present investigation of this thesis. Thus, a number of cases

[363] Andrew P. Simester, John R. Spencer, Findlay Stark, G. R. Sullivan and Graham J. Virgo, *Simester and Sullivan's Criminal Law: Theory and Doctrine* (7th ed. Hart Publishing 2019).
[364] *Ibid*., [5.5(ii)].
[365] Andrew Ashworth and Jeremy Horder, *Principles of Criminal Law* (7th ed. Oxford University Press 2013), 181 – 186.
[366] Horder (2019), 205 – 209.
[367] Alan Norrie, *Crime, Reason and History: A Critical Introduction to Criminal Law* (3rd ed. Cambridge University Press 2014), 82 – 88.
[368] Tatjana Hörnle, 'Social expectations in the criminal law: The "reasonable person" in a comparative perspective' (2008) 11(1) *New Criminal Law Review: An International and Interdisciplinary Journal* 1, 23 – 28.
[369] Horder (2019), 205 (emphasis added).

are considered in this section to demonstrate how certain concepts in negligent offences would operate under the hybrid approach.

#### *10.5.2.1. Simpson v Peat*

The quintessential and common example of an offence that may be committed by negligence is that of careless and inconsiderate driving (or driving without due care and attention).[370] The defendant in *Simpson v Peat*[371] was convicted for driving without due care and attention following a collision with a motorcyclist. Both had been driving at a reasonable speed in opposite directions when the defendant, in what was described by the Magistrate at trial as an "error of judgment", attempted to turn onto a minor road leading off to the right, crossing the path of the motorcyclist and causing the collision. The appeal (to the High Court by way of case stated) substantively concerned whether it was possible for an error of judgment to support a conviction for driving without due care and attention, in light of preceding authority[372] and the argument that a driver must have been paying attention in order to make a judgment, even if that judgment was wrong in the event.[373]

The High Court considered that an error of judgment could indeed be negligent, notwithstanding that attention had been paid in the circumstances. The key factor was whether or not the defendant's conduct fell below the requisite standard of care:

> '[W]as the defendant exercising that degree of care and attention that a reasonable and prudent driver would exercise in the circumstances? If he was not they should convict; if, on the other hand, the circumstances show that his conduct was not inconsistent with that of a reasonably prudent driver, the case has not been proved.'[374]

This is a classic exposition of the reasonable man test which pervades both the civil and criminal law of negligence, and is retained in the objective first limb of the hybrid test.

[370] Road Traffic Act 1988, s. 3.
[371] *Simpson v Peat* [1952] 2 QB 24.
[372] *R v Howell* (1938) 27 Cr App R 5.
[373] *Simpson* [1952], 26.
[374] *Ibid.*, 27 – 28.

Clearly, therefore, it is possible for an error of judgment to fall below the requisite standard of care if a reasonably careful and prudent driver would not have made that same error in the circumstances. Equally, however, the judgment closed by affirming that the fact that an accident or injury has occurred does not necessarily mean that an individual has fallen below the requisite standard of care. The Court offered the hypothetical of a driver who, in an emergency and through no fault of his own, swerves right to avert a collision but ends up causing another accident, whilst it is shown that the accident would not have occurred had he swerved left:

> '[T]hat is being wise after the event and, if the driver was in fact exercising the degree of care and attention which a reasonably prudent driver would exercise, he ought not to be convicted, even though another and perhaps more highly skilled driver would have acted differently.'[375]

On the available facts, the High Court was in no doubt that the defendant's conviction was safe; he had cut across the line of traffic coming in the opposite direction in order to make the turn; the motorcyclist was not found to have been driving at unreasonable speed; the defendant had also failed to signal; and, in any event, it was his responsibility to take sufficient care to ensure that he could execute such a manoeuvre safely.[376]

A similar finding would be inevitable upon the application of the hybrid approach to negligence. Regarding the first limb, the same objective standard of the reasonable man is applied to determine whether or not the defendant's conduct was objectively unreasonable which, on the aforementioned facts of *Simpson*, was evidently the case. This having been found, it is presumed that it is perfectly reasonable to expect anybody in the same circumstances to appreciate that such a manoeuvre – turning across oncoming traffic without signalling – was unreasonable. No evidence was offered in order to rebut this presumption and so the defendant would similarly be convicted.

[375] *Ibid.*, 28.
[376] *Ibid.*, 28 – 29.

#### *10.5.2.2. R v Bannister*

Whereas the test for reasonableness in negligence is unassailably an objective one adjudged by the standards of the hypothetical reasonable person, both legislation and the courts have explored a number of circumstances and characteristics subjective to individual defendants which may nonetheless have bearing on the application of that objective test. Remaining with driving offences, the offences of causing death by dangerous driving,[377] dangerous driving,[378] and careless or inconsiderate driving[379] may each be committed where the defendant drives negligently – *i.e.,* below (for careless driving) or far below (for dangerous driving) the standard expected of a reasonably competent and careful driver. Meanwhile, for the purposes of ascertaining what standard is expected of a competent and careful driver, regard is given not only to those circumstances of which a defendant is expected to be aware 'but also to any circumstances shown to have been within the knowledge of the accused.'[380] The case of *R v Bannister*[381] concerned whether or not a defendant's superior driving skill and qualifications were such circumstances for which regard should be given, and how such regard interacts with the ostensibly objective reasonable man test.

The defendant in *Bannister* was an experienced road traffic police officer who had completed advanced driving courses teaching him to drive at very high speeds. On the occasion resulting in his conviction, it was 18:00 when the roads were liable to be busy, it was dark on account of being January, and the defendant accelerated on the motorway from approximately 88 to 120 miles per hour during torrential rain and with a lot of surface water on the road. At one point the car slid and spun out of control, aquaplaning across the road and into a copse of trees at the roadside. Although there were no injuries caused, the defendant's car was written off and he was convicted of dangerous driving. On appeal, the defendant contended that his superior driving qualifications amounted to

[377] Road Traffic Act 1988, s. 1.
[378] *Ibid.*, s. 2.
[379] *Ibid.*, s. 3.
[380] *Ibid.*, ss. 2A(3) & 3ZA(3).
[381] *R v Bannister* [2009] EWCA Crim 1571.

knowledge that ought to be taken into consideration when applying the reasonable man test and considering whether or not he had driven without due care and attention.

The Court of Appeal was mindful of, and considered itself to be bound by, its own previous decisions where matters such as a defendant's liability to lose consciousness during hypoglycaemic episodes,[382] consumption of cocaine,[383] and consumption of alcohol[384] were considered relevant to the question of whether or not driving had been objectively dangerous. However, the Court rejected that such factors had any impact upon the objective test of the reasonably competent and careful driver; rather, they are:

> '[F]acts relating to the condition of the driver which are as relevant as the driver's knowledge of the unroadworthiness of a car or the conditions of the weather on the road. Those facts can be taken into account without in any way departing from the test of the competent and careful driver.'[385]

In this regard, the defendant who knows that they may suffer a hypoglycaemic attack whilst driving, or is under the influence of cocaine or alcohol, is still assessed against the objective standard of care, those factors clearly suggesting that the standard has been breached when the individual decided to drive in spite of them. Conversely, the same cannot be said in relation to an individual's attested superior driving skills. Where the relevant test is that of the reasonably competent and careful driver, to take consideration of some superior or special skill would necessarily change the test into that of the especially skilled driver, which would be a fundamentally different test to apply. The defendant's conviction for dangerous driving was commuted to careless driving on the technical ground that the judge had incorrectly directed the jury.

[382] *R v Marison* [1997] RTR 457.
[383] *R v Pleydell* [2005] EWCA Crim 1447.
[384] *R v Woodward* [1995] 2 Cr App R 388.
[385] *Bannister* [2009], [18].

The hybrid approach to negligence would clearly support a conviction for dangerous driving in *Bannister*.[386] Concerning the first limb of the test, driving at 120 miles per hour in the dark and torrential rain, and at a time when the roads are likely to be busy, undoubtedly falls far below the standard of the reasonably competent and careful driver. Further, regarding the additional requirement of obviousness for the offence of dangerous driving, driving in the manner described creates an objectively obvious risk of endangering life or property. With regards to the second limb of the test, the defendant's appeal to have their superior driving qualifications taken into consideration would not assist in this case. Whilst such qualifications might reflect the defendant's increased driving skill, they ought also to reflect a heightened awareness of the risks and dangers of driving in the conditions that are described. Consequently, in answering the question whether or not it is reasonable to expect anybody in the same circumstances to appreciate the unreasonableness of their driving, the defendant's superior training might in fact be argued to lend greater support to an affirmative answer. Certainly, in any event, the defendant's additional driving qualifications are unlikely to be regarded as sufficient to rebut the presumed reasonable expectation that anybody in the same circumstances would appreciate the objective unreasonableness of driving in the conditions described.

#### *10.5.2.3. R v Price and Bell*

Moving away from negligent driving offences, there has been some degree of inconsistency with regards to other offences and whether or not subjective characteristics of the defendant might be taken into consideration in the application of the objective reasonable man test. The case of *R v Price and Bell*[387] related to the prosecution of a number of serving members of the army for the negligent performance of duty.[388] In brief, the soldiers had been conducting a training exercise using general purpose machine guns with live ammunition, with Price responsible for overseeing the exercise. At some point, Bell's gun jammed with a live round in the chamber unbeknownst to him; whilst attempting to fix the issue, he placed the gun on the floor pointing towards another soldier.

[386] More serious than careless driving, dangerous driving not only requires that the defendant's standard of care fell *far* below that of the reasonable competent and careful driver, but also that it would be obvious to the reasonable driver that driving in that way would be dangerous to life or property.
[387] *R v Price and Bell* [2014] EWCA Crim 229.
[388] Armed Forces Act 2006, s. 15(2).

In investigating the jam, Bell caused an unintended discharged which killed another, and he pleaded guilty at trial to negligently performing his duty whilst handling the weapon. Price pleaded not guilty to the same charge, which was brought on the grounds that he did not properly supervise the soldiers regarding guns jamming, nor did he determine the cause of the jam nor the safety of the weapon, nor did he prevent the weapon from being moved from its firing point thereby endangering others, and he failed to direct Bell regarding unjamming the weapon safely.

As circumstances giving rise to his appeal, Price pleaded that he had worked predominantly as a clerk and then financial systems administrator and had not undertaken weapons training for some 14 years before being sent on a tour of duty to Kenya. There, he was instructed that he would be responsible as a safety supervisor at the firing range, to which he raised concerns with his Captain regarding his lack of experience for the role. The defendant then underwent a brief training session which included basic handling skills of the general-purpose machine gun; although he did not fire the gun himself and was not fully aware of all the faults which could materialise with the weapon, he passed the gun's handling test. The defendant's key grounds of appeal were that the trial judge in his summing up had misdirected that the defendant's lack of training and experience for the task of firing range safety supervisor was immaterial to the judgment of negligence and, furthermore, that the judge had similarly failed to include these deficiencies in the application of the reasonable man test.

The Court of Appeal in *Price and Bell* was mindful of previous authority such as *Bannister*, discussed above,[389] providing that a defendant's superior or, indeed, inferior skill in a particular matter is not a relevant consideration in the application of the reasonable man test. However, the Court was equally mindful that Price had not held himself out as possessing any skill and had in fact registered his misgivings about the appointment as safety supervisor to his superior officer. The Court thus found that, 'in the somewhat special circumstances of the service context' the requisite standard of care would be that which is 'expected of the reasonable serviceman *having similar training,*

[389] Also, *R v Bateman* (1927) 19 Cr App R 8, 12.

*knowledge and experience* as the accused.'[390] That notwithstanding, the Court considered that there were a range of obvious and common-sense precautions that the defendant might nonetheless have taken notwithstanding his lack of experience and training, such as halting the exercise once one of the machine guns became jammed, and ensuring that the weapon was never pointing towards others whilst the jam was investigated. Consequently, the defendant's conviction was upheld.

A similar finding is readily supportable under the hybrid approach to negligence, albeit the defendant's inferior training and experience is dealt with slightly differently. Regarding the first limb of the test, it is clear that the litany of basic failures on behalf of Price were objectively unreasonable; at the very least, the training exercise ought to have been halted so the weapon could be properly inspected, whilst at no point whatsoever should the weapon have been pointing towards another person. It is under the second limb of the test that the defendant's inferior skill, training and experience with the weapons becomes a relevant consideration in asking whether it is reasonable to expect anybody with similarly inferior skill, training and experience to appreciate that their conduct fell below the standard expected of reasonable man.

In this respect, the defendant attracts a certain degree of sympathy as he had both expressed his own concern at being the safety supervisor and accepted that he was not aware of all of the issues that could develop with the weapon. Nonetheless, for the same reasons expressed by the Court of Appeal, even with the lack of training and experience it is reasonable to expect somebody to appreciate the basic failures of not stopping the exercise, investigating the jammed weapon safely, and at all times ensuring that it was not pointing towards others. These precautions are both obvious and not dependent upon any superior training with weapons; therefore, the presumption under the second limb of the test is not rebutted, and the hybrid approach to negligence similarly supports the defendant's conviction in *Price and Bell*, notwithstanding the relevance of the defendant's inferior experience and skill.

[390] *Price and Bell* [2014], [20].

#### 10.5.2.4. *R (on the application of the RSPCA) v C*

Whereas the decision in *Price and Bell* might be limited to the military context within which that case took place, another characteristic that the courts appear minded to take into consideration in crimes of negligence is the defendant's age and intellect.[391] The defendant in *R (on the application of the RSPCA) v C*[392] was the 15-year-old owner of a cat who lived at home with her father, whom the Magistrates' Court accepted was the joint owner of the animal. At some time, the cat sustained an injury to its tail which went untreated for an unknown period of time, perhaps two weeks according to available estimates. Both the girl and her father were prosecuted for causing unnecessary suffering to an animal;[393] the father pleaded guilty, who had apparently decided that the cat did not require veterinary treatment unless its condition worsened, and his conviction formed one of the findings of fact for the daughter's prosecution.

The question on appeal to the High Court, by way of case stated, was the relevance of the daughter's age and subservience to her father in the reasonableness of her decision not to seek further care for the injured cat. The High Court concluded that the Magistrates were entitled to find that defendant's acquiescence to her father's decision was reasonable, owing to the fact of the defendant's young age and that she had 'relied on the decision and actions of her father.' This was, in the circumstances, considered to be an entirely reasonable thing for a 15-year-old girl to do, as no doubt countless teenagers rely at some point on the advice of an adult parent. Thus, the defendant had not acted negligently and her acquittal at trial was upheld.

As with *Price and Bell*, above, the hybrid approach to negligence in *R (on the application of the RSPCA)* would reach the same substantive conclusions, but gives a slightly different treatment to the defendant's subjective characteristics. The first limb of the test remains purely objective and, on the facts of the case, both the Magistrate's and the High Court were minded that the failure to seek veterinary attention had been unreasonable in the circumstances. The defendant's age and subservience to her father become relevant

[391] In this respect, see also *R v Hudson* [1966] 1 QB 448, 455.
[392] *R (on the application of the RSPCA) v C* [2006] EWHC 1069.
[393] Protection of Animals Act 1911, s. 1(1)(a).

considerations under the second limb of the hybrid test which, it is recalled, takes the form of a rebuttable presumption in cases of negligence. Thus, the question is whether or not those subjective characteristics of the defendant would be sufficient to rebut the presumption that it is reasonable to expect anybody in the same circumstances to appreciate the unreasonableness of their conduct.

In concurrence with the Magistrates' Court and High Court, it is submitted that the age, maturity and subservient familial position of a child may be patently relevant factors which rebut the aforementioned presumption. Children and teenagers both lack the same insight and foresight as adults in circumstances such as those in the present case and, furthermore, it is not only reasonable but expected that such individuals will take direction from their parents, as did the defendant in the present case. It is therefore readily arguable that the presumption is rebutted in this case, and it is not reasonable to expect anybody in the same circumstances as the defendant to appreciate that their conduct had been unreasonable in the event. This example is therefore useful for demonstrating both how subjective characteristics remain relevant in offences of negligence and, furthermore, how they may be used to rebut the presumption under the second limb of the hybrid test.

#### 10.5.2.5. *R v Colohan*

The aforementioned authorities notwithstanding, it is important to emphasise that the orthodox and most widespread approach to criminal negligence is to apply a purely objective reasonable man test which does not take the defendant's subjective characteristics into consideration. This is roundly demonstrated in the case of *R v Colohan,*[394] which concerned the prosecution for harassment[395] of a defendant who had sent a number of letters to his local Member of Parliament. The letters were mostly rambling and incoherent concerning a number of real and imaginary issues; however, they also contained a certain amount of abuse and material that could be construed as threatening violence or death towards the individual. The defendant neither answered questions during police interview nor gave evidence at trial, but medical evidence was

[394] *R v Colohan* [2001] EWCA Crim 1251.
[395] Protection from Harassment Act 1997, ss. 1 & 2.

provided in his defence which stated that he was suffering from schizophrenia at all material times which resulted in disordered and obsessive thoughts, delusions, and beliefs in different conspiracies. The doctor's evidence was that the defendant's letters were a product of his schizophrenia, albeit there was no suggestion that the defendant's illness reached the high threshold of an insanity defence, nor that he was unaware of the fact that what he was doing was wrong.[396]

The defendant was duly convicted, and the question on appeal was whether the fact of his schizophrenia was relevant to the test of whether a reasonable person in possession of the same information would consider that the course of conduct amounted to harassment.[397] The Court of Appeal determined that this factor could not be taken into consideration in the application of the reasonable man test as to do so would be to judge the defendant instead 'by the standards of the hypothetical reasonable schizophrenic',[398] and the defendant's conviction was consequently upheld. Whilst this affirms an orthodox application of the reasonable man test, this decision could also be read as being somewhat restricted to the specific facts of the case and, in particular, the historical background to the Protection from Harassment Act 1997. The decision can be read as one highly governed by considerations of policy and, therefore, potentially limited to the offence of harassment.[399]

One of the principal aims behind the Act was to protect people from stalking, an offence which is often likely to be committed by 'those of obsessive or otherwise unusual psychological make-up and very frequently by those suffering from an identifiable mental illness.'[400] The argument follows, to dilute the reasonable man test with consideration of a defendant's mental illness risks undermining the very purpose of the legislation, which is to both protect the victims of stalking (including from people whose stalking is driven

[396] See *Loake v Director of Public Prosecutions* [2018] QB 998, 1013.
[397] Protection from Harassment Act 1997, s. 1(2).
[398] *Colohan* [2001], [10] & [20] – [21].
[399] David Ormerod, 'Trial: Direction to jury – Reasonable person – Reasonable conduct – Defendant suffering from paranoid schizophrenia' (2001) (Oct) *Criminal Law Review* 845, 846.
[400] *Colohan* [2001], [18].

by mental illness) whilst 'also increase[ing] the chances of the offender receiving psychiatric assessment and treatment.'[401]

The hybrid test of negligence would likely take a different approach to the Court of Appeal in *Colohan*, although reaching a similar conclusion. There is little difficulty with the objective limb of the test as sending numerous letters to another containing threats of violence is undoubtedly an unreasonable course of conduct to take. With regards to the second limb of the hybrid test, the defendant's schizophrenia is a relevant consideration to the question of whether or not it is reasonable to expect anybody to appreciate that their behaviour is objectively unreasonable. On the one hand, the medical evidence went so far as to suggest that the letters were a product of the defendant's illness in the respect that he both 'believed unshakably' the things he wrote and he 'would have felt compelled to write them.'[402] What is not stated, however, is whether the defendant 'was denying that he knew what he was doing in writing the letters, denying that he knew that the letters constituted a course of harassing conduct, or was claiming simply that the harassment was reasonable.'[403]

Certainly, as the defendant declined to give evidence, the defence did not adduce any evidence as to whether or not his illness was such as to impact upon a person's capacity to appreciate that their conduct was objectively unreasonable. As such, it is reasonable to argue that the presumption under the second limb of the hybrid test has not been rebutted, even by evidence of the defendant's schizophrenia, and his conviction would be similarly upheld. Alternatively, were the defendant's schizophrenia sufficient to amount to a defence, it is submitted that the court would be justified in returning a verdict of "not responsible" – this is a novel suggestion that is explored further in section 12.3.2, below. The consequence of a not responsible verdict would be that the defendant would neither obtain a criminal record nor be subject to more punitive elements of punishment for the purposes of general or specific deterrence, or expressivism. However, the court would remain empowered to impose rehabilitation through a hospital order to ensure that the

[401] *Ibid*.
[402] *Ibid*., [6].
[403] Ormerod (2001), 846.

defendant received appropriate treatment for their mental illness, as well as incapacitation as necessary to secure that treatment and the wider safety of society, and restitution to the victim as appropriate.

#### *10.5.2.6. R v Adomako*

The facts of this case are stated in section 10.2.2.10 of this thesis, above, and shall not be rehearsed here in full; briefly, the case concerned the conviction of an anaesthetist for involuntary manslaughter in circumstances where a vital connection for a patient's ventilator had come loose during surgery and had not been checked in the ensuing minutes as the patient's oxygen levels deteriorated. In other crimes of negligence considered thus far, the term negligence is perhaps used imprecisely where some element of the offence is subject to the standard of the reasonable person, such as in the offence of harassment where the element regarding which a defendant may be negligent is knowledge that a reasonable person would possess concerning a course of conduct amounting to harassment. Negligence is approached more precisely akin to the civil law concept in the offence of gross negligence manslaughter, which was the substantive offence at issue in *R v Adomako*.

The requisite components of negligence *simpliciter* are well established in civil law and, being composed entirely of objective elements, do not require particular elaboration in this thesis. Stated briefly, the defendant must first have owed the victim a duty of care recognised in law; such duties are developed incrementally and must usually be identified within established categories already recognised in jurisprudence. Where no such precedent can be found, the courts may nonetheless develop new legal duties of care in circumstances where harm was a reasonably foreseeable consequence of the defendant's actions, there existed a relationship of proximity between the parties, and it is fair, just and reasonable in the circumstances to impose such a duty.[404] The foreseeability of harm is arguably the main test here, as there will be practically very few circumstances in which

[404] *Caparo Industries plc v Dickman* [1990] 2 AC 605, 617 – 618.

death will have been reasonably foreseeable without there also being a relationship of proximity.[405]

Second, there must be a breach of that standard of care, measured according to the hypothetical reasonable person test discussed throughout this section. Third, the defendant's breach of that duty of care must have been the actual cause of the victim's death, applying the ordinary rules of causation in negligence.[406] In addition to these ordinary principles of negligence, the offence of gross negligence manslaughter requires two more elements; fourth, the breach of duty in question must have given rise to a serious and obvious risk of causing death to another. The relevant test is whether the 'reasonable prudent person possessed of the information known to the defendant would have foreseen that the defendant's actions or omissions constituting the breach of duty had exposed the deceased to an "obvious and serious" risk of death', applying an objective test infused with certain characteristics subjective to the defendant.[407] Fifth and finally, the defendant's conduct must have been grossly negligent, which is to say that:

> '[It] went beyond a mere matter of compensation between subjects and showed such disregard for the life and safety of others, as to amount to a crime against the State and conduct deserving of punishment.'[408]

The expert medical evidence in *Adomako* attested that the defendant's conduct in treating his patient had been 'abysmal' and amounted to a 'gross dereliction of care.'[409] The defendant was therefore convicted, and his appeals were dismissed on the basis that the trial judge had correctly directed the jury to consider the case through the lens of gross negligence.[410]

---

[405] Jonathan Herring and Elaine Palser, 'The duty of care in gross negligence manslaughter' (2007) (Jan) *Criminal Law Review* 24, 30; citing *Alcock v Chief Constable of South Yorkshire Police* [1992] 1 AC 310, 396; *Perrett v Collins* [1998] 2 Lloyds Rep 255, 262.
[406] *Inner South London Coroner, ex parte Douglas-Williams* [1999] 1 All ER 344, 350.
[407] *R v Kuddus* [2019] EWCA Crim 837, [35].
[408] *Bateman* (1927), 11 – 12.
[409] *Adomako* [1995], 182.
[410] See further Richard Card and Jill Molloy, *Card, Cross & Jones Criminal Law* (22nd ed. Oxford University Press 2016), 266.

Only minor amendments are required to the hybrid test for negligence in order to align it with the higher requirements of gross negligence manslaughter. First, the objective limb of the test is expanded to require not only that the defendant's conduct is unreasonable, but that it is *grossly* unreasonable, applying the same extant approach whereby gross negligence is simply a matter of degree. The second limb of the test proceeds to reflect the fourth element of gross negligence, described above, except this is also expanded so as to ask whether it is reasonable to expect anybody in the same circumstances to appreciate the obvious and serious risk of death arising from a breach of the duty concerned. This expands the subjective circumstances that may be taken into consideration beyond merely the defendant's knowledge under the extant test; however, in keeping with the broader approach to hybrid negligence, this second limb operates by way of a rebuttable presumption.

Applying this hybrid approach to gross negligence to the facts of *Adomako*, the defendant's conviction is similarly supported. Considering the first limb of the test in light of the expert evidence, the defendant's failure to check something as basic as the patient's ventilation tube was evidently grossly unreasonable in the circumstances. The expert evidence attested that the hypothetical reasonable anaesthetist would have checked something so obvious. With regards to the second limb of the test, it is self-evident from the facts of the case that the failure to check a ventilation tube during surgery would create a serious and obvious risk to the patient's life. The defendant submitted that 'after things went wrong I think I did panic a bit';[411] however, people trained and holding themselves out to be anaesthetists must be expected to keep a degree of composure during the difficulties of surgery, and it is unlikely that "a bit" of panic would be sufficient to rebut the presumption under the second limb of the test. Stated in full, it is reasonable to expect that any anaesthetist under the pressure of an emergency situation would nonetheless appreciate the serious and obvious risk of death from failing to ensure that a ventilation tube was properly connected when a patient's oxygen levels were falling. The presumption is not rebutted, and the defendant is guilty of gross negligence manslaughter.

[411] *Adomako* [1995], 182.

### *10.5.3. Final Comments on Negligence*

Criminal liability satisfied by negligence has received notable academic criticism over the years, most particularly in relation to the absence of any requirement to demonstrate a defendant's culpable state of mind. For example, Alexander, Ferzan and Morse[412] contest the argument that negligence may be culpable because of a person's failure to advert to a risk that they had a fair chance to perceive (had they so tried), even though they did not consciously choose to create an unreasonable risk of harm:

> 'We disagree. The world is full of risks to which we are oblivious. Or, more accurately, because risk is an epistemic, not ontic, notion, (meaning that it is one contingent upon knowledge rather than one based upon factual existence) we frequently believe we are creating a certain level of risk when someone in an epistemically superior position to ours would assess the risk to be higher or lower than we have estimated. Sometimes, the epistemically superior position is the product of better information: for example the doctor knows that what we believe is just a mole is in fact a life-threatening melanoma. At other times, we have failed to notice something that another might have noticed, or we have forgotten something that another might have remembered…
>
> 'We are not morally culpable for taking risks of which we are unaware. At any point in time we are failing to notice a great many things, we have forgotten a great many things, and we are misinformed or uninformed about many things. An injunction to notice, remember, and be fully informed about anything that bears on risks to others is an injunction no human being can comply with, so violating this injunction reflects no moral defect.'[413]

[412] Larry Alexander, Kimberly Kessler Ferzan and Stephen Morse, *Crime and Culpability: A Theory of Criminal Law* (Cambridge University Press 2009).
[413] *Ibid*., 70.

It is respectfully submitted that there are a number of flaws in this argument; most glaring being the absence of any reference to reasonableness. Even under the extant law of negligent criminal offences, individuals are not liable to be responsible for each and every potential risk their actions may pose to others; nor are they liable for each and every one of those risks that materialises in the event. Firstly, crimes of negligence are necessarily restricted by their *actus reus*; it is only those risks of producing a prohibited *actus reus* to which an individual needs be cogniscent. Secondly, it is inherent within negligence that not every breach of a risk is liable, but only those where the defendant's conduct fell below the standard of the ordinary, reasonable man. In these respects, it is submitted that Alexander, Ferzan and Morse inflate the onerous burden of offences committed by negligence.

Furthermore, and notwithstanding the general rejection of subjective states of mind as a basis for legal responsibility supported throughout this thesis, it is proposed that the hybrid approach to negligence may settle this debate in relation to *culpability*. Specifically, the hybrid approach to *mens rea* holds individuals responsible not because a specific *state of mind* is proven, but because it is demonstrated that the mind was possessed of three capacities which make it culpable. The first two capacities – that the mind is responsive to reason and exercises ordinary self-control – are presumed in law to exist for all adults but may be rebutted by particular defences. The third capacity is that which is encapsulated by the second limb of the hybrid test, that is the capacity for the defendant to appreciate the nature and consequences of their actions as they relate to the particular *mens rea* of an offence – *e.g.*, the capacity to appreciate that their conduct breaches standards of reasonableness in offences of negligence.

If it is accepted that culpability is founded in the existence of these three mental capacities (whether exercised or not at the time of an offence), then the hybrid approach to negligence offences addresses the claim that such offences lack proof of culpability. The fact that the third capacity (like the first two) becomes a rebuttable presumption in relation to negligent offences neither eradicates the fact that it is a requisite component of the offence, nor disables the defendant in pleading relevant subjective circumstances and

characteristics in order to support certain defences.[414] Concurrently, by treating the second limb of the hybrid test as a rebuttable presumption, the *status quo* of the extant law is maintained insofar as the prosecution are not required to prove any *mens rea* in relation to negligent offences, but may still be required to respond to certain positive defences raised by the defendant.

It is notable that the hybrid approach to attaching culpability to conduct performed inadvertently is broadly similar to that proposed by H. L. A. Hart.[415] Hart rejected the imposition of an entirely objective and impersonal standard of conduct for offences of negligence, instead suggesting that responsibility might follow if two questions can be answered affirmatively. First, '*did the accused fail to take those precautions which any reasonable man with normal capacities would in the circumstances have taken?*' This is virtually equivalent to the first limb of the hybrid test asking whether or not that defendant's conduct was unreasonable – *i.e.*, has he failed to take those precautions that a reasonable person would. Second, *'could the accused, given his mental and physical capacities, have taken those precautions?'* Again, this is eminently similar to the second limb of the hybrid test which takes into consideration the defendant's subjective capacities in asking whether or not it was reasonable to expect them to appreciate the unreasonableness of their conduct.

[414] Albeit, not all defences are available to offences of negligence and strict liability.
[415] H. L. A. Hart, 'Negligence, m*ens rea*, and criminal responsibility' in Hart H. L. A. (ed.), *Punishment and Responsibility: Essays in the Philosophy of Law* (2nd ed. Oxford University Press 2008), 154.

# 11. Defences

> '[A]lthough responsibility is assessed in a social context, the capacity to learn social norms and the capacity to act in accordance with them are matters of individual brain function. It is precisely because an important difference exists between a normal brain and the brain of someone who is seriously demented or unreachably deluded that such people are not considered responsible for crimes they might commit. Moreover, judicial institutions rely on threat of punishment to deter. The late maturation of the prefrontal cortex (with reference to neuronal density, synaptic density, dendritic length and myelination) means that the brains of mature adults are critically different from those of young children – which almost certainly accounts for the child's more modest ability to appreciate the consequences of his or her choices and to resist temptation.'
>
> - Patricia S. Churchland, 2005.[1]

As has been alluded to throughout the previous chapter of this thesis, it is proposed that legal defences can fundamentally be related to one or more of the three mental capacities the existence of which, it is argued, are prerequisites for imposing legal responsibility for a person's actions. That is to say, defences operate to negate legal responsibility because they suggest that one or more of these three capacities has been sufficiently diminished or abrogated entirely. Taking the defence of temporary insanity (by reason of a defendant's schizophrenia) as a hypothetical example, it is well recognised that schizophrenia and other psychiatric disorders can have the effect of rendering a person incapable (or, at least, significantly less capable) of recognising and incorporating reason into their decision-making, can diminish a person's ability to exercise ordinary self-control over their actions, and can impede an individual from appreciating the nature of

[1] Patricia S. Churchland, 'Brain-base values' (2005) 93(4) *American Scientist* 356, 357 – 358.

their actions and the consequences they may have in the world. Whilst this represents perhaps the most complete example of how a legal defence may interfere with the three capacities required for responsibility, it is proposed that all defences can be related in this way to at least one (but potentially two or even all three) of the aforementioned mental capacities.

## *11.1. Justifications and Excuses*

An orthodox description of defences in English criminal law describes two categories: defences that justify an action following which 'we accept responsibility but deny that it was bad'; and defences that excuse an action whereby 'we admit that it was bad but don't accept full, or even any responsibility.'[2] Expressing the distinction more fully, Baron writes:

> '[T]o say that an action is justified is to say… that though the action is of a type that is usually wrong, in these circumstances it was not wrong. To say that an action is excused, by contrast, is to say that it was indeed wrong (and the agent did commit the act we are saying was wrong), but the agent is not blameworthy.'[3]

Duff explains the argument that where a justificatory defence applies then, considering all matters on balance, the relevant action was at least permissible if not entirely justifiable. For example, breaking the window of a burning building to rescue somebody trapped inside is obvious justified in the circumstances. Where an excusatory defence is applied, the relevant action remains unjustified and condemnable; however, there are 'features of the action's context or of the agent given which it would be unjust or unfair' to attribute blame.[4] One such hypothetical example might the breaking of a window during an epileptic fit. However, the distinction between justificatory and excusatory defences has developed considerably without underlying principle, and there is often little clear

[2] John L. Austin, 'A plea for excuses' (1956-57) 57 *Proceedings of the Aristotelian Society* 1, 2.
[3] Marcia Baron, 'Justifications and excuses' (2005) 2(2) *Ohio State Journal of Criminal Law* 387, 389 – 390.
[4] R. Anthony Duff, *Answering for Crime: Responsibility and Liability in the Criminal Law* (Hart Publishing 2007), 265.

dividing line between the two.[5] As Duff writes, 'one might indeed suggest that the orthodox distinction between justifications and excuses is now so misleading, given the assumption that it provides an exhaustive classification of defences, that we should abandon the terms altogether.'[6]

The distinction between justificatory and excusatory defences arguably has considerably less importance today than in previous times. Perhaps most starkly, until 1828 a defendant who was acquitted of homicide following an excusatory defence was still liable to have their goods forfeited by the crown; although such a killing was not regarded as a crime, it was nonetheless 'universally regarded as deplorable' and was, therefore, not devoid of consequences.[7] Since then, however, whilst there may be some limited implications of the justification / excuse distinction – for example, with regards to whether a justified action may be resisted in civil law – 'the same rationale leads in either case to blameless acquittal… [and] *there is no fundamental difference of output in the criminal law*, at least for individual defendants.'[8] Nor does the justification / excuse dichotomy provide any particular guidance in what gives defences their exculpatory character or, indeed, how new defences ought to be identified and developed. Not least owing to the unprincipled development of the law itself in this area, it is difficult to identify precisely what circumstances or characteristics should or should not be permitted to excuse an individual defendant's otherwise criminal conduct.

## *11.2. Defences and Capacities*

Whilst important in order to appreciate the extant law, the present thesis does not propose to expand further upon the justification / excuse dichotomy; certainly this distinction, which no longer has any significant *practical* impact in the treatment of individual defendants, equally has no particular relevance to the theory of responsibility defended in

[5] For example, see Kent Greenawalt, 'The perplexing borders of justification and excuse' (1984) 84(8) *Columbia Law Review* 1897.
[6] Duff (2007), 265.
[7] John C. Smith, *Justification and Excuse in the Criminal Law* (Stevens & Sons Ltd. 1989), 7.
[8] Andrew P. Simester, 'On justifications and excuses' in Zedner L. and Roberts J. (eds.), *Principles and Values in Criminal Law and Criminal Justice: Essays in Honour of Andrew Ashworth* (Oxford University Press 2012), 96 (emphasis added); see also Smith (1989), 7.

this work. Rather, it is submitted that all defences may be regarded within the wider concept of *mens rea* through the manner in which defences address one of more of the three crucial capacities. The two previous chapters of this thesis have argued that *mens rea* is broadly comprised of three capacities. The first two capacities – that an individual is responsive to reasons and possesses ordinary self-control – encapsulate the requirement that action is voluntary and so attaches to the *actus reus* of an offence as a rebuttable presumption. The third capacity – that an individual appreciates the nature of their conduct as it relates to a specific form of objectively defined *mens rea* – is essentially what is at issue within the hybrid conception of *mens rea*.

As the previous chapter of this thesis has extensively demonstrated, it must first be proven that the defendant's conduct matches the relevant *mens rea* of an offence as objectively defined; for example, that causing death or serious injury was a virtually certain consequence of their actions (*i.e.*, intention) for the offence of murder. Second, it must be proven that it was reasonable to expect anybody in the defendant's subjective circumstances to appreciate the nature and consequences of their conduct, *i.e.*, to appreciate the virtual certainty of causing death or serious injury by their actions for the offence of murder. In resisting the prosecution on the second limb of this test, the defence will introduce extraneous circumstances and endogenous characteristics which suggest that it is *not* reasonable to expect anybody in the same position to so appreciate the nature of their actions. Those factors that are relevant for consideration – and, indeed, which therefore limits the matters that a defendant may seek to raise – are those that impact upon the two presumed capacities and the final capacity which must be proven for an individual to appreciate the nature and consequences of their conduct.

In turn, it is submitted that defences operate to the effect that one or more of these three crucial capacities underlying *mens rea* have been sufficiently diminished or abrogated entirely; that is to say, that capacity is operating at a level so deficiently below that of the reasonable person that it is no longer reasonable to hold the individual (entirely) responsible for their decisions and actions. For example, the violent attacker who pleads that their actions were the direct result of paranoid delusions arising from their schizophrenia may raise an insanity defence. If successful, such a defence could speak to

all three of the relevant aforementioned mental capacities: the effects of schizophrenia can indeed render an individual's mind and decision-making capacities unresponsive to normal reason and argument; it can impact upon their ability to control certain behaviours and bodily actions; and it can emphatically impact upon a person's ability to appreciate the nature of their actions. In this sense, the legal defence of insanity, which may be proven on facts resulting from an illness such as schizophrenia, clearly has the potential to touch upon all three of the mental capacities which underlie the concept of *mens rea*.

However, humanity and society are diverse and the hypothetical reasonable person encompasses all manner of behaviour and conduct that might fairly be regarded as reasonable, even if contrary to the particular moral standards of certain individuals or groups. Thus, it is submitted, subjective circumstances and characteristics that rise to the level of undermining one or more of the aforementioned capacities must do so to such a degree that the capacity in question is diminished beyond the bounds of the ordinary reasonable man. In this sense, whereas something like schizophrenia can readily be appreciated for having a significant, virtually inescapable (without treatment), and distorting effect upon a person's mental capacities, mere personality traits, quirks or characteristics *alone* are unlikely to amount to such a degree as to sufficiently overwhelm or undermine the mental capacities underlying *mens rea*. It is for this reason, for example, that an individual simply being quick to temper will not generally support a legal defence.

The reasoning follows that everybody, including the hypothetical reasonable person, experiences anger and temper at times in life whilst, inevitably, some are quicker to anger than others. But it is equally within the bounds of ordinary, reasonable behaviour for people to exert self-control and contain their emotional outbursts, at least to the extent that they do not descend into criminality. Thus the law, supported by ample medical evidence, accepts that a condition such as schizophrenia may overwhelm those ordinary capacities of self-control, for example, whilst merely being an individual who is quicker to anger is not regarded as sufficient to overwhelm the reasonable capacities of ordinary self-control that are shared by and expected of all ordinary people. One important implication of this relatively high threshold for undermining the three capacities in regular neurotypical adults is that legal defences should be relatively limited in number and

reasonably slow to establish anew. Whilst the final section 11.4 of this thesis proposes a new defence of addiction that may be implicated and developed from the present theory of responsibility, the expectation is that established legal defences will already cover a large proportion of the situations where circumstances and characteristics have been found to meet such a sufficiently high threshold so as to undermine the three capacities underlying *mens rea*.

It may fairly be asked whether or not this proposed rationale behind legal defences – *i.e.,* as circumstances or conditions which sufficiently diminish, undermine, or entirely abrogate one or more of the three mental capacities that are crucial for responsibility – is fairly applicable to all defences. For example, it might be contested that certain defences such as self-defence are recognised not because of some diminution of mental capacities but more simply because people have a right to defend themselves against attack. In the first instance, there is little difficulty in applying the proposed reasoning to those defences traditionally categorised as excuses, namely mistake, intoxication, insanity, automatism, diminished responsibility, loss of control and duress. One reason is that these defences do not assert that the defendant has acted in any way legitimately, or according to any right or privilege; rather, the traditional excusatory defences eponymously excuse an individual for having done something that is nonetheless recognised as wrong. This is because, second, the majority of excusatory defences are premised upon characteristics of the individual defendant that are well recognised as significantly reducing one or more of their relevant capacities: insanity and diminished responsibility require proof of certain recognised mental conditions or defects of reasoning; automatism and loss of control respectively require the full or partial loss of ordinary self-control over bodily actions; and the manner in which intoxication can seriously inhibit people's ability to think rationally and control their actions scarcely requires rehearsing.

Duress (by threat or circumstances) is perhaps anomalous in this regard,[9] as the defence does not require that the individual defendant is suffering from some mental deficiency

[9] Mistake is also somewhat anomalous insofar as a mistake of fact does not provide a defence *per se*, but is relied upon to either deny the existence of *mens rea* in the first place or to establish another affirmative defence – see further section 11.3.2, below.

or characteristic that is well recognised for diminishing one or more of their relevant capacities. However, an examination of this defence (conducted more thoroughly in section 11.3.8 of this thesis, below) nonetheless reveals that the defence is indeed concerned with relevant capacities underlying decisions to act. In brief, the defence can only be established in response to threats of immediate death or serious injury from another, or impersonal and exogenous circumstances giving rise to the same, and where the defendant has no opportunity to escape that threat and seek protection from lawful sources. Crucially, the relevant threats must be of such seriousness, immediacy and gravity that any other 'sober person of reasonable firmness, sharing the characteristics of the defendant,' would have acted in the same way in response to those threats.[10]

From its earliest conceptions,[11] duress has been recognised as a defence because even the hypothetical reasonable man is not expected to resist the threat of death or serious injury to himself or his loved ones; the voluntariness of any person is diminished in the face of such threats.[12] Where the other excusatory defences point to *characteristics of the particular defendant* which diminish their relevant capacities such as responding to reason or exercising self-control, duress points to the *characteristics of the particular threat* against which those same capacities of *any* defendant would reasonably be expected to diminish. In this regard, as with the other excusatory defences, duress does not claim that the response of the defendant in submitting to a given threat was justified or that they enjoyed any right to commit a criminal act, but that their actions were excused by the nature and impact of that threat.

Regarding those defences traditionally identified as justificatory – principally self-defence and necessity – orthodoxy explains that the defendant does not commit a wrongful act at all. That is to say, in the circumstances in which either defence is successfully pleaded, the defendant's actions were the correct one's to perform, or they possessed the right to so act as they did. Whether or not the law recognises any such *right*

[10] *R v Graham* (1982) 74 Cr App R 235, 241.
[11] See Amy Elkington, 'The historical development of duress and the unfounded result of denying duress as a defence to murder' (2022) 0(0) [online] *Journal of Criminal Law* 1.
[12] *R v Hudson and Taylor* [1971] 2 QB 202, 206; John Hyman, *Action, Knowledge, and Will* (Oxford University Press 2015), Ch. 4; Dennis Patterson, 'Rethinking duress' (2016) 7(3) *Jurisprudence* 672.

to act in an otherwise criminal manner under these defences is not specifically contested in the present thesis. What is submitted is that these defences can also be understood in relation to how relevant circumstances impact upon one or more of a defendant's crucial capacities for responsibility. As Uniacke writes, even if it accepted that defendants acting in self-defence do so under a degree of compulsion, 'this would not necessarily preclude the use of force in self-defence being objectively justified conduct.'[13] To the extent that the law consequently endorses any such positive right for a defendant to act in an otherwise criminal manner within the context of these justificatory defences, it is further submitted that that right to act has been recognised secondarily and because of the primary impact that these defences have on those crucial capacities.

Following the above discussion on duress, the defence of self-defence is readily appreciated in relation to the concept of voluntariness and the underlying capacities of responding to reason and exercising ordinary self-control. Self-defence is conceptually very similar to duress as both are fundamentally concerned with how a defendant responds to immediate threats of violence; 'the motivational factor in each case is precisely the same.'[14] Perhaps the most important distinction, however, is that the defendant acting in self-defence seeks to protect against the threat of violence typically by responding with their own violence against the issuer of that threat; in duress, however, the defendant acquiesces to the threat of violence typically at the expense of an innocent third party.[15] Indeed, this distinction likely explains why self-defence is available against any (threat of) unlawful violence and may also defend against the charge of murder, whereas duress is only available against the threat of death or serious injury and provides no defence against murder. Crucially, whereas self-defence might be conceived as providing a *right* to respond to threats of unlawful violence with violence of one's own, it is submitted that, like duress, self-defence is underpinned by the recognised impact that threats of violence have upon *any* person's volition.

[13] Suzanne Uniacke, *Permissible Killing: The Self-Defence Justification of Homicide* (Cambridge University Press 1994), 35.
[14] Warren Brookbanks, 'Compulsion and self-defence' (1990) 20(1) *Victoria University of Wellington Law Review* 95, 96.
[15] Uniacke (1994), 30 – 31.

Discussed further in section 11.3.7, below, Blackstone describes self-defence as 'immediate justice to which [the defendant] is prompted by nature, and which *no prudential motives are strong enough to restrain*.'[16] As Sangero writes more recently, the 'common situation' which activates each of the defences of duress, self-defence, and necessity (addressed below) is circumstances of compulsion – 'immediate danger to a certain legitimate interest that *forces* the actor to harm another interest in order to save the first.'[17] He adds further, '[i]n the context of the difficult situation in which the actor finds himself, *the survival instinct acts very powerfully*. This is human nature.'[18] Whilst Sangero proceeds to identify further features of self-defence which distinguish it from other defences of compulsion and permit the use of this defence against the charge of murder, for example, the touchstone that activates this defence remains with the threatening circumstances which *compel* the defendant to act in defence of their interests.

Further still, considering that the recipient of responsive violence under self-defence is the issuer of the original unlawful violence, this defence most ostensibly arises in circumstances of "fight-or-flight" when the defendant must flee the scene or otherwise has no realistic choice but to defend their legitimate interests. The immediacy of the threatened or actual unlawful violence is a further necessary condition of the defence which precludes much opportunity to escape or seek law enforcement, in which case only responsive violence remains. Thus, as with duress, the voluntariness of the defendant is significantly diminished within the context of self-defence, and with it one or both of the capacities for responding to reason and exercising ordinary self-control. That self-defence may also provide a positive *right* for the defendant to respond with their own reasonable violence is a potential additional feature of this defence, but does not alter its genesis in compulsion and the diminution of voluntariness.

Necessity is the other prominent legal defence that is traditionally categorised as being justificatory, discussed in section 11.3.8, below. In essence, the claim follows that when an individual is faced with an impossible choice between two harmful outcomes, they

[16] William Blackstone, *Commentaries on the Laws of England in Four Books, Volume 2* (George Sharswood (ed.), J. B. Lippincott Co. 1875), 2 – 4 (emphasis added).
[17] Boaz Sangero, *Self-Defence in Criminal Law* (Hart Publishing 2006), 2.
[18] *Ibid.* 37 (emphasis added).

may justifiably select the lesser of two evils even if this would otherwise ordinarily constitute a criminal offence. As with self-defence, the traditional dichotomy between justification and excuse suggests that a defendant acting out of necessity is not being excused for some personal characteristic which diminishes their capacities, but that their chosen action was legitimate and correct in the circumstances, despite normally constituting a criminal offence. However, further similarly to self-defence (and duress), it is submitted that "circumstances of compulsion" again provide the touchstone for initiating this defence, in which circumstances the voluntariness of the defendant is sufficiently undermined by the absence of any "real" choice at all. In this regard, Blackstone referred to "inevitable necessity" as a 'species of defect of will' and a 'constraint upon the will, whereby a man is urged to do that which his judgment disapproves.'[19] Again, this does not preclude the possibility that the law may elevate actions taken out of necessity to the status of some form of legal *right* to act; only that, underlying the finding of any such right, the fundamental basis for the defence lies in the impact that the circumstances of compulsion have upon the defendant's (and, indeed, any reasonable person's) voluntariness, consisting of the capacities to respond to reason and exercise ordinary self-control.

This conception of necessity has been developed most fully in Canada where the Supreme Court considers decisions made out of necessity as being "morally involuntary." Young explains, a morally involuntary act 'is not involuntary in the sense that actions of a person in a condition of automatism are involuntary, but they are *morally* involuntary because the actor had *no reasonable alternative* to breaking the law. The act was wrongful, but it was an *acceptable failure brought on by normal human weakness*.'[20] Again drawing comparisons between necessity, duress and self-defence, these defences of compulsion do not claim that the capacities of any individual defendant have been diminished *per se* due to some mental illness or deficiency particular to them, but that the relevant circumstances of compulsion *would be sufficient to override the voluntariness of any*

[19] William Blackstone, *Commentaries on the Laws of England in Four Books, Volume 4* (Edward Christian (ed.), A. Strahan and W. Woodfall 1795), 27; see further Walter H. Hitchler, 'Necessity as a defence in criminal cases' (1929) 33(3) *Dickinson Law Review* 138, 141 – 144.
[20] Diana Young, 'Excuses and intelligibility in criminal law' (2004) 79(1) *University of New Brunswick Law Journal* 79, 97 (emphasis added).

*reasonable person placed in the same circumstances*. Additionally, it is submitted that one of the central purposes of punishment – *i.e.,* enforcing a criminal prohibition and deterring against its breach – is critically undermined in necessity (and other defences of compulsion), because the punishment threatened tends to be less onerous in the moment than the alternative of breaching the law.[21]

Thus, the seminal Canadian case of *R v Perka*[22] concerned a ship carrying cannabis to Alaska but which was forced by a storm to make port in Canada and repair damage; the crew were subsequently charged with the illegal import of drugs and pleaded necessity. Dickson CJ considered that the underlying premise of the defence of necessity is involuntariness: '[r]ealistically… his act is not a "voluntary" one. His "choice" to break the law is no true choice at all; *it is remorselessly compelled by normal human instincts*.'[23] Dickson CJ offered the hypothetical example of a mountaineer who is faced with the "choice" of breaking into a cabin in the mountains or freezing to death. Whilst a person in such circumstances does have a choice in the literal sense of the word – they could technically physically restrain themselves from committing the criminal act:

> '[R]ealistically, his choice is not a "voluntary" one… this was a limited choice, constrained and indeed created only as a result of the existing circumstances. It is accordingly a choice forced upon the offender.'[24]

Although it is accepted that moral involuntariness does not amount to the same as literal involuntariness such as in the case of an automaton, it is nonetheless readily appreciable how even the most reasonable man has no real option but to submit to criminality in the relevant circumstances giving rise to necessity, duress or self-defence. Whilst referring to circumstances of compulsion as involving a diminution of volition might be regarded as "stretching" the concept, it is submitted that the greater stretch would be to consider to

[21] For example, see Hitchler (1929), 140.
[22] *R v Perka* [1984] 2 SCR 232; see further *R v Ruzic* [2001] SCR 687; *R v Ryan* [2013] SCC 3; Steve Coughlan, 'The rise and fall of duress: How duress changed necessity before being excluded by self-defence' (2013) 39(1) *Queen's Law Journal* 83.
[23] *Ibid.,* 249 (emphasis added).
[24] Glenys Williams, 'Necessity: Duress of circumstances of moral involuntariness?' (2014) 43(1) *Common Law World Review* 1, 8.

such circumstances as entirely or ordinarily volitional choices. Reiterating the crucial argument, it is not submitted that the capacities of the individual defendant are abnormally diminished, but that circumstances of compulsion sufficiently reduce *any reasonable person's* volition such as to abrogate responsibility for their subsequent actions.

Finally, it is at this stage plain to appreciate why defences are considered within the broader concept of *mens rea*, because both are effectively concerned with the existence and operation of the three aforementioned mental capacities. It is submitted that this conceptualisation is more helpful today than the justification / excuse dichotomy for two key reasons; it explains both why legal defences (at least partially) extinguish responsibility for actions and, consequently, how new legal defences can be recognised and developed. The explanation for both is that defences undermine or overwhelm one or more of the capacities for reasons responsiveness, ordinary self-control and the appreciation of the nature of conduct. Because one or more of these capacities is no longer functioning as is accepted as being necessary for an individual to properly comply with the criminal law, they are not regarded as being responsible for their conduct. Equally, where circumstances and characteristics can be shown to overwhelm or undermine one or more of these capacities, they may form the basis of developing a new legal defence. This latter point is explored at the end of the present chapter with the proposal for a partial defence of addiction, recognising in particular how addiction can severely impact upon the presumed capacities for reasons responsiveness and ordinary self-control.

## 11.3. Testing Hybrid Defences

### 11.3.1. Bare Denial of Mens Rea

The first defence to consider is not strictly a "defence" within either the extant justification / excuse dichotomy nor the proposed reconceptualization of defences by reference to crucial mental capacities. Nonetheless important to consider for completeness, a bare denial of *mens rea* amounts to the statement that the prosecution have not exercised their burden of proof with regards to the *mens rea* of the offence charged. Recalling the burden of the prosecution to prove the facts of their case beyond reasonable doubt, this means that both limbs of the hybrid approach to *mens rea* must be

proven to this same standard. Where the defence therefore argues a bare denial of *mens rea,* they are stating that the prosecution has failed to prove one or both of two things: that the defendant's conduct fits within the requisite *mens rea* for the offence as objectively described; and / or that it is reasonable to expect anybody in the same circumstances as the defendant to appreciate how their conduct relates to that *mens rea* element of the offence. It is reiterated, the initial burden at trial ordinarily falls upon the prosecution to prove both of these elements beyond a reasonable doubt; a bare denial of *mens rea* asserts that there exists reasonable doubt with regards to one or both of these elements and that the prosecution have therefore failed to execute their burden.

Examples where a bare denial of *mens rea* might have been argued with regards to the objective first limb of the hybrid test include the cases of *Hyam v Director of Public Prosecutions*[25] and *R v Nedrick.*[26] Proceeding along virtually identical facts, both cases concerned defendants who had set fire to the homes of their rivals by pouring accelerant through the front-door letterbox at night when they knew that the properties were occupied and the occupants were likely to be sleeping. Both defendants denied any intention to kill or cause grievous bodily harm but asserted that they only wished to frighten their intended victim whereas, in both events, others who had been sleeping in the properties succumbed to the fires. Interestingly, both defendants were convicted of murder at trial but, whereas both also appealed, only the defendant in *Nedrick* was successful in having their conviction quashed. Whilst the appeal was undoubtedly successful upon technical grounds, the Court of Appeal substituted the conviction for one of manslaughter and maintained the original sentence of 15 years' imprisonment. Nonetheless, these two cases decided upon similar facts reach different conclusions upon whether the dangerous and wholly unjustifiable conduct so described amounted to murder.

Regarding the objective first limb of the hybrid test for intention, it was argued that setting a house fire in the circumstances described is indeed virtually certain to result in death or serious injury to occupants known to be sleeping inside. House fires are exceedingly dangerous by their very nature; the fires were set when occupants were likely to be

[25] *Hyam v Director of Public Prosecutions* [1975] AC 55.
[26] *R v Nedrick* [1986] 1 WLR 1025.

sleeping; the fires were set blocking one of the main exits from each property; and no warning was given nor emergency services called. Even if death could not be said to be virtually certain from such circumstances, it is readily arguable the serious injury is virtually certain, whether from inhaling toxic fumes, oxygen deprivation, suffering burns, or further injuries obtained whilst attempting to escape. That being said, it is also at least arguable that the undoubted danger of housefires nonetheless fails to reach the near-inevitability of being virtually certain. The aforementioned dangers notwithstanding, there were 52 fatalities and 1,014 casualties out of approximately 3,667 deliberately started housefires in England in 2018/19, representing a mortality rate of around 1.4% and a casualty rate of 27.7%,[27] which arguably may not amount to virtual certainty.

Then again, the circumstances of the individual cases must be taken into account, and it might be argued that the dangers were inherently increased in *Hyam* and *Nedrick* by starting the fires at night, blocking a key route of exit, and ensuring that the property was occupied at the time. Ultimately, as is the case with *mens rea* generally, the decision of whether certain conduct fits within an objective description of the requisite *mens rea* of an offence is a matter to be determined by the jury, applying the knowledge and experiences that are common amongst the snapshot of ordinary society that the jury represents. Cases such as *Hyam* and *Nedrick* offer contentious examples where the objective limb of the hybrid test might reasonably be argued either way.

With regards to the second limb of the hybrid test, a bare denial of *mens rea* here argues that the prosecution has failed to prove that it is reasonable to expect anybody in the same circumstances as the defendant to appreciate the nature of their conduct as it relates to the particular type of *mens rea* for the offence in question. For example, in the case of *R v Hancock and Shankland*[28] considered in section 10.1.2.5 of this thesis, above, the defendants pushed concrete blocks from a bridge and into the path of a taxi. Their purported intention was only to frighten the occupant of the taxi, but in the event the

[27] Home Office, 'FIRE0402: Fatalities and non-fatal casualties in deliberate fires by fire and rescue authority' (Home Office Fire Statistics Data Tables, November 2020) <https://www.gov.uk/government/statistical-data-sets/fire-statistics-data-tables#deliberate-fires-attended> accessed 12/01/2021.

[28] *R v Hancock and Shankland* [1986] AC 455.

blocks hit the vehicle windscreen and killed the driver. The defendants were convicted of murder but successfully appealed, and their conviction was substituted for manslaughter. The analysis of this case under the hybrid test for intention proved difficult with regards to the second limb in particular, and the only relevant subjective evidence submitted by the defence was that their intention had been only to frighten to occupant of the taxi and that they had therefore aimed their projectiles in front of, but not directly at, the vehicle.

The difficulty in this case arises because the relevant charge was that of murder; therefore, under the second limb of the hybrid test, it would be necessary to prove that it was reasonable to expect anybody who was aiming a projectile at the road in front of a moving vehicle to appreciate that death or serious injury was virtually certain to result from their actions. This could be answered affirmatively – the act of pushing concrete blocks from a bridge into the path of moving traffic is so clearly and inherently dangerous that the fact that the individual was not aiming directly at a vehicle should not excuse any reasonable person from failing to appreciate the virtual certainty of death or serious injury following from their actions. Then again, the defendants were aiming away from a small moving target, at distance and using a heavy and cumbersome projectile, and it could be argued that any other person similarly aiming not to hit a difficult moving target would not reasonably be expected to appreciate the virtual certain of causing death or serious injury. Again, this is a line which the jury will need to decide, and for which the principal burden falls upon the prosecution to prove in their favour beyond any reasonable doubt.

It is crucially important to recall that the second limb of the hybrid approach to *mens rea* takes into consideration any relevant circumstances or characteristics subjective to the defendant and which potentially have a bearing on the three capacities underlying *mens rea*. Thus, the defendant introduces such relevant factors in order to persuade the jury that it is *not* reasonable to expect anybody in their same circumstances to have appreciated the nature of their conduct as it related to the *mens rea* of the offence charged. The cases of *R v Bello*[29] and *Elliott v C*[30] exemplify this point; *Bello* concerned a defendant accused of knowingly remaining in the UK beyond the time permitted on their visa, whilst *Elliott*

[29] *R v Bello* (1978) 67 Cr App R 288.
[30] *Elliott v C* [1983] 1 WLR 939.

concerned a defendant charged with recklessly causing criminal damage in relation to a fire they started to keep warm. Whilst the first limb of the hybrid tests for knowledge and recklessness were satisfied in each case respectively, factors subjective to each defendant were raised which might suggest that it would be unreasonable to expect anybody in the same circumstances to appreciate the nature of their actions.

In *Bello*, the defendant claimed that they had forgotten (and, therefore, did not possess) knowledge of the date of expiry for their visa as a result of grief following the death of their mother. The defendant had been nonetheless capable of carrying on other normal activities such as attending polytechnic college, and it was determined that ordinary grief alone is insufficient to absolve an individual of responsibilty for forgetting criminally relevant knowledge. In *Elliott*, conversely, the defendant's young age and maturity were clearly sufficient factors which were demonstrably in play at the time of the defendant's offending and, therefore, it was ultimately unreasonable to expect anybody in the same circumstances to appreciate the unreasonable risk of their actions. *Bello* thus presents an example where potentially relevant subjective characteristics of the defendant are taken into consideration but deemed nonetheless insufficient to render it unreasonable to expect that anybody in the same circumstance to would appreciate the nature of their actions. Meanwhile, *Elliott* demonstrates a likely successful application of the bare denial of *mens rea* "defence" as, on the evidence provided, it was unreasonable to expect anybody of the defendant's young age and immaturity to have necessarily appreciated the unreasonable risks of setting a fire in the circumstances of that case.[31]

### *11.3.2. Mistake*

As with a bare denial of *mens rea*, mistake is not a legal "defence" in the strict sense of the word; rather, a defendant may claim that they have made a mistake in order to establish another legal defence or, indeed, in order to deny that they possessed the requisite *mens rea*.[32] For example, if a woman picks up another's identical handbag

[31] See also *R v Clarke* [1972] 1 All ER 219, in which a moment of absent-mindedness induced by diabetic depression following sugar deficiency did not amount to a "defect of reason" for the purposes of an insanity defence, but instead supported the defence of an absence of *mens rea*.
[32] Jonathan Herring, *Criminal Law: Text, Cases, and Materials* (9th ed. Oxford University Press 2020), 705.

mistaking it to be her own, she does not possess the intention to deprive another or their property and nor is she acting dishonestly, by virtue of the fact that she is operating under a material mistake which negates the requisite *mens rea*. Alternatively, if one man sees another raise his arms in the air and mistakenly believes that he is about to be attacked, that mistake may form the basis of the fear of immediate violence that is required to establish the defence of self-defence. Thus, the fact of having made a mistake does not itself absolve an individual of responsibility; that mistake must operate to negate *mens rea* or provide the foundations of a positive defence such as self-defence or necessity.

A crucial distinction between the operation of a mistake in negativing *mens rea* and in establishing another positive defence lies in whether or not that mistake must be reasonable. The orthodox position in common law previously held that an 'honest and reasonable belief in the existence of circumstance which, if true, would make the act for which a prisoner is indicted an innocent act has always been held to be a good defence,'[33] imposing an objective test of whether any given mistake was a reasonable one to make in the circumstances.[34] The House of Lords overturned this orthodoxy almost a century later in *Director of Public Prosecutions v Morgan*,[35] in a case which proceeded along shocking facts. Four defendants were drinking and decided to seek out a sex worker. Upon being unable to find one, however, one defendant told his friends that they could all go to his house and sleep with his somewhat estranged wife; he added that whilst she might appear to be resisting, this would be an act and that she gained a sexual thrill from playing out a rape fantasy. The defendants arrived at the home, grabbed the victim and took her to the bedroom and then each proceeded to rape her; the victim protested and fought during the entire event but was held down and overcome by the defendants. After the event, the victim presented herself to hospital and reported the attack.

The critical question before the House of Lords was whether or not the defendants' mistaken belief in the victim's consent could provide a defence to the charge of rape if that belief was genuinely held, but otherwise regardless of how unreasonable that belief

[33] *R v Tolson* (1889) 23 QBD 168, 181.
[34] See also *Bowman v Blyth* (1856) 7 E&B 26, 43; *R v Prince* (1875) LR 2 CCR 154, 175.
[35] *Director of Public Prosecutions v Morgan* [1976] AC 182.

might have been. By a majority of three-to-two the Court determined the question in the affirmative, not because a genuinely held belief provides a defence *per se* but, because an individual could not be said to intentionally raping another – *i.e.*, intentionally having *non-consensual* sex with another – if they genuinely believed that the other person was consenting. This was a 'matter of inexorable logic… either the prosecution proves that the accused had the required intent, or it does not.'[36] However, the House of Lords provided more generally that there exists a distinction between raising mistake as a negation of *mens rea* and raising mistake as the basis of a defence, and with regards to the latter any such mistake must be reasonable.[37] Lord Cross provided:

> 'If the words defining an offence prove either expressly or impliedly that a man is not to be guilty of it if he believes something to be true, then he cannot be found guilty if the jury think that he may have believed it to be true, however inadequate his reasons for doing so. But, if the definition of the offence is on the face of it "absolute" and the defendant is seeking to escape his *prima facie* liability by a defence of mistaken belief, I can see no hardship to him in requiring the mistake – if it is to afford him a defence – to be based on reasonable grounds.'[38]

As may be appreciated, the decision in *Morgan* switched the orthodox position with regards to mistakes negating *mens rea* specifically, determining the test to be entirely subjective to the defendant making a mistake honestly, even if entirely unreasonably. Whilst this general approach to mistake continues to hold today,[39] the reformed law of rape under the Sexual Offences Act 2003 explicitly departs from *Morgan* in requiring that a defendant's belief in consent be reasonable for important policy reasons, not least the importance of obtaining consent in contrast with the limited imposition that such a requirement places upon any potential defendant.[40] In addition, whereas any mistake

---

[36] *Ibid*., 214.
[37] *Ibid*., 214 & 238; see Glanville Williams, *Criminal Law: The General Part* (Stevens & Sons Ltd. 1953), 163 & 167.
[38] *Ibid*., 202 – 203.
[39] Confirmed in *R v Kimber* [1981] 3 All ER 84; *R v Beckford* [1988] AC 130; *Director of Public Prosecutions v B* [2000] 2 AC 428; *R v K* [2001] UKHL 41.
[40] See further Home Office, *Protecting the Public: Strengthening Protection Against Sex Offenders and Reforming the Law on Sexual Offences* (Cmnd 5668, 2002), [32] – [34]; Jennifer Temkin and Andrew

providing a foundation for a positive defence must generally be reasonable, there exists an erroneous exception in relation to self-defence discussed in section 11.3.7 of this thesis, below. It is also important to highlight that, notwithstanding their successful appeal on the point of law, the defendants' convictions in *Morgan* were upheld by the House of Lords as, on the evidence available, no reasonable jury could have found that the defendants reasonably believed in the victim's consent.

The decision in *Morgan* has understandably garnered considerable criticism concerning some of the perspectives and conclusions regarding rape which are undoubtedly informed by the time when that decision was taken in the 1970s. For one, the decision was denounced as a "rapist's charter" by those who considered that unscrupulous defendants would take advantage of merely having to demonstrate an honest mistake, however unreasonable, in order to escape conviction. However, this criticism overlooks that fact that the House of Lords determined to maintain the defendants' convictions notwithstanding their successful appeal. McAuley explains:

> '[T]he "rapist's charter" argument overlooks that fact that a defendant who says he honestly but unreasonably believed that his victim was consenting will not succeed *unless the jury thinks there is a reasonable chance that his story might be true*. Because of the essentially ascriptive character of belief, juries will inevitably ask themselves whether they would have believed what the defendant claims to have believed had they been in his shoes; and the more implausible his story appears when set against the facts as a whole, the less likely they are to believe it… In this sense, *Morgan* does not disturb the traditional requirement that mistakes must be reasonable; it merely entrusts it to juries in the guise of an informal criterion for assessing the validity of mistakes.'[41]

Ashworth, 'The Sexual Offences Act 2003: (1) Rape, sexual assaults and the problems of consent' (2004) (May) *Criminal Law Review* 328, 340 – 341; Jenny McEwan, '"I thought she consented": Defeat of the rape shield or the defence that shall not run?' (2006) (Nov) *Criminal Law Review* 969.

[41] Finbarr McAuley, 'The grammar of mistake in criminal law' (1996) 31 *Irish Jurist* 56, 71.

The "inexorable logic" of distinguishing between the two varieties of mistake is more clearly illuminated by Simester.[42] He argues, when mistake is raised for the purpose of supporting a positive defence such as self-defence, duress or necessity, the defendant knows that they are committing a *prima facie* wrong. That is to say, the individual who punches another person in defence, or partakes in a robbery under duress, or breaks the speed limit to rush somebody to hospital, is committing an offence advertently (*i.e.,* not unknowingly, accidentally, or unintentionally). Consequently:

> '[He] recognizes that he is inflicting harm, and knows that his actions require justification… [the defendant] is asserting a liberty, based upon circumstances, to inflict harm *knowingly*: it does not seem too much to ask for reasonable ascertainment of such circumstances.'[43]

In stark contrast, when mistake is raised for the purpose of negating *mens rea* with regards to a particular element of the offence, the defendant is unaware that they are doing anything *prima facie* wrong. So, neither the woman who mistakenly picks up somebody else's bag that is identical to her own, nor the man who jovially slaps a stranger on the back mistakenly believing it to be his brother, has any reason to suspect that they are stealing another's property or assaulting a stranger respectively, and so they lack the requisite *mens rea* for either offence. Simester contrasts the hypothetical examples of a person who shoots their wife believing her to be a rabbit and another person who shoots their wife in the belief that they are a burglar launching an attack. He writes:

> 'My killing another person is *prima facie* wrongful, and I know that I ought to be very sure of having good reasons for doing so before I embark upon such a course of action. It is quite a different matter when what I think I am doing is not *prima facie* wrongful: I need no reasons for my conduct.'[44]

[42] Andrew P. Simester, 'Mistakes in defence' (1992) 12(2) *Oxford Journal of Legal Studies* 295.
[43] *Ibid*., 309.
[44] *Ibid*., 311.

It follows that to require mistakes negating *mens rea* to be explicitly reasonable would offend the inexorable logic of how mistakes and *mens rea* interact, whereas to require that mistakes substantiating defences are reasonable goes no further than 'requiring a citizen to take reasonable care to ascertain the facts relevant to his avoiding doing a prohibited act.'[45]

*

The approach to both mistakes negating *mens rea* and mistakes supporting legal defences is readily subsumed within the general hybrid approach to *mens rea* adopted throughout this thesis and, furthermore, can be linked to the capacity for appreciating the nature and consequences of one's conduct. Beginning with the former type, simply put, a mistake that potentially negates *mens rea* is included within the relevant circumstances that a defendant may submit under the second limb of the hybrid test. For example, where a defendant picks up another's bag mistaking it for their own, the fact that they subjectively made that mistake is a relevant consideration under the second limb of the test (assuming that they convince the jury that they did, indeed, genuinely hold that mistake). Moreover, so too are the facts and circumstances surrounding and supporting that mistake relevant for consideration; for example, the fact that the defendant's own bag looks identical or, indeed, completely different to the victim's would be eminently relevant. Thus, for the *mens rea* of intention in the offence of theft, the question under the second limb of the hybrid test becomes, *is it reasonable to expect anybody mistakenly taking another's identical / non-identical bag to appreciate the virtual certainty that they are appropriating another's property?*

By including both the mistake itself and the relevant facts supporting / undermining that mistake under the second limb of the hybrid test, that mistake does in fact fall to be assessed against its surrounding facts and whether, taken together, they support or undermine the reasonable expectation that anybody would appreciate the nature of their conduct. In the aforementioned hypothetical example, therefore, if the appropriated bag is entirely different to the defendant's, the second limb of the hybrid test must be

[45] *Sweet v Parsley* [1970] AC 132, 165.

answered in the affirmative. No matter how honest the mistake, when any reasonable person sees that their bag has miraculously changed appearance, it is reasonable to expect that they would appreciate the fact that they have taken somebody else's property, regardless of the purported honesty of their mistake. Conversely, if the hypothetical defendant has mistakenly picked up another's bag that appears identical to their own, it is hardly reasonable to expect that they would appreciate the *virtual certainty* of having stolen another's property, having regard to *both* the fact of their mistake *and the fact that the circumstances support that mistake*.

Following this approach, it is plain to see how even the defendants' purportedly *honest* mistakes regarding consent in *Morgan* would not be likely to successfully negate *mens rea* under the hybrid approach. The defendants' assertion is that they mistakenly but genuinely believed in the victim's consent; the circumstances in their favour are the apparent permission given by the victim's husband along with his explanation that she would pretend to object in order to obtain a sexual thrill from the experience. The circumstances against the defendants include the facts that the defendants had neither obtained explicit consent from the victim, nor discussed the plan with her, nor even met her; and that the victim loudly, forcibly and continually protested and fought against the entire attack.

Applying the second limb of the hybrid test for belief (in consent), *is it reasonable to expect anybody in the circumstances described to appreciate the absence of a conviction in consent?* Inexorably, the answer is yes; even if the defendants honestly believed the husband's assertions, the lack of any prior contact with the victim, of any explicit or even implied consent, and the violent protestations throughout the attack, all serve to undermine the reasonableness of a belief in consent based upon even a genuine mistake for the purposes of the hybrid test. That is, notwithstanding the purported honesty of the defendants' mistaken belief, it remained nonetheless reasonable to expect that anybody in the same circumstances would appreciate the lack of conviction in consent.

Where mistake is raised by the defendant in support of a positive defence, the correct approach follows the extant law in requiring that mistake itself to be based upon

reasonable grounds. Thus, for example, the defendant may raise their mistaken perception of a threat of violence or grave circumstances in order to support a defence of duress or necessity respectively. Owing to the requirement that such a mistake be based on reasonable grounds, however, the defendant will need to offer such grounds in support of their alleged mistake, as opposed to merely relying on their own testimony that the mistake was genuinely held. Meanwhile, the prosecution will be able to defeat the entire defence by proving either that the claimed mistake was unreasonable in the circumstances or not genuinely held; *i.e.*, beyond reasonable doubt the grounds did not exist to support the assertion that the mistake was reasonable (or, indeed, genuine). This follows the current law whereby a claimed mistake founding a legal defence must itself be reasonable, in contrast to a mistake negating *mens rea* which is considered amongst the defendant's subjective circumstances for the purposes of the second limb of the hybrid *mens rea* test.

### *11.3.3. Intoxication*

The relevance of intoxication poses a notable challenge within the existing approach to *mens rea* and legal culpability. On the one hand, it is trite that intoxication can have a range of powerful impacts on brain function and resultant behaviour: intoxication may render a person less likely to foresee certain (even obvious) risks and, equally, more likely to take risks (whether or not they have been foreseen); intoxication can render people less effective at interpreting scenarios or the behaviour of others and, equally again, less capable of appreciating the effects of their own actions; and intoxication can render people more impulsive and less deliberative in their decisions, and can exacerbate certain emotional responses such as by rendering people more aggressive or confrontational, to offer some examples.

On the other hand, there are strong reasons why it might be considered unreasonable for the law to place too much countenance on intoxication as providing a defence to criminal acts. In contrast to other defences such as insanity, automatism, or even duress and self-defence, intoxication is notably common with virtually all adults experiencing the effects of alcohol at least once in their lifetime. Further, with the effects of intoxication being so well documented, an expansive intoxication defence risks providing a *carte blanche* for

criminal conduct, with prospective defendants being able to escape liability simply by carrying out their criminal intentions under the influence of drugs and / or alcohol.

In balancing these contentions, therefore, the law arrives at a position informed predominantly by public policy, 'plac[ing] the public interest in preventing and combatting alcohol- and drug-related crime over the deontological question of whether an individual can actually be properly blamed for actions committed in an intoxicated state.'[46] By way of brief overview, a defendant's *involuntary* intoxication may be adduced to raise reasonable doubt concerning their *mens rea* for offences of both basic and specific intent, explored further below, and provides a complete defence where successfully pleaded. Conversely, *voluntary* intoxication can only be relevant to offences of specific intent, whilst providing no possible defence to offences of basic intent. Moreover, a successful plea of voluntary intoxication will only provide a partial defence to offences of specific intent, reducing a conviction to its lesser basic intent equivalent; for example, reducing a charge of murder to manslaughter.

Notably, intoxication itself does not provide a general defence *per se*, but is adduced in order to deny that the defendant possessed the requisite *mens rea* for the offence charged. The leading authority on the subject is provided by the House of Lords in *Director of Public Prosecutions v Majewski*,[47] concerning a defendant charged with assault occasioning actual bodily harm and assaulting a police officer. His defence consisted of the claim that he had "completely blacked out" following the consumption of large quantities of alcohol and drugs.

> 'In so far as culpability is concerned, the law has always drawn a distinction between the man who voluntarily makes himself drunk and cases of involuntary intoxication (*e.g.,* where a person's drink is unknowingly to him doctored by another at a party). … In *R v Pearson*

[46] Michael Bohlander, 'From *Marx* to *Majewski*: A review of the law on voluntary intoxication in the former German Democratic Republic' in Livings B., Reed A. and Wake N. (eds.), *Mental Condition Defences and the Criminal Justice System: Perspective from Law and Medicine* (Cambridge Scholars Publishing 2015), 276.

[47] *Director of Public Prosecutions v Majewski* [1977] AC 443.

> (1835) 2 Lew. 144 it was said that if a party be made drunk by the stratagem or fraud of another, he is not responsible for his actions. Accordingly, drunkenness *may be taken into consideration to explain his conduct*. … If a man of his own volition takes a substance which causes him to cast off the restraints of reason and conscience, no wrong is done to him by holding him answerable criminally for any injury he may do while in that condition. His course of conduct in reducing himself by drugs and drink to that condition in my view supplies the evidence of *mens rea*, of guilty mind certainly sufficient for crimes of basic intent. It is a reckless course of conduct and recklessness is enough to constitute the necessary *mens rea* in assault cases: see *R v Venna* [1976] QB 421, per James LJ at p. 429. The drunkenness is itself an intrinsic, an integral part of the crime, the other part being the evidence of the unlawful use of force against the victim. Together they add up to criminal recklessness.[48]

A number of concepts may now be unpacked from this *dictum*. First is the distinction between offences of basic and specific intent, which was formulated in three ways in *Majewski*.[49] Lord Elwyn-Jones suggested that the *mens rea* for crimes of basic intent 'does not go beyond the *actus reus*' whereas the defendant must intend both the specific *actus reus* and some further consequence of that *actus reus* in crimes of specific intent.[50] For example, where the (basic intent) offence of assault may be committed with an intention to do that act which causes another to apprehend immediate and unlawful violence, the (specific intent) offence of wounding with intent to cause grievous bodily harm requires an intention to make the bodily actions which cause another's wounding *and* an intention that such wounding is actually caused by those actions. In a similar vein, Lord Simon distinguished the two types of offence according to the purposive intention which must be proven for crimes of specific intent, *i.e.,* proof of intention of the consequences of a particular *actus reus*.[51] Third, each of Lords Elwyn-Jones, Simon and

[48] *Ibid.,* 462 – 463 & 474 – 475.
[49] See further Alan R. Ward, 'Making some sense of self-induced intoxication' (1986) 45(2) *Cambridge Law Journal* 247.
[50] *Majewski* [1977], 471; citing *Morgan* [1976], 216.
[51] *Ibid.,* 479; citing *R v George* (1960) 128 Can CC 289, 301.

Russell referred to crimes of specific intent as being those that can only be committed intentionally, whereas crimes of basic intent can be committed recklessly;[52] and it is this latter distinction that has largely become the settled approach.[53]

A second question arising from *Majewski* concerns the precise interaction between voluntary intoxication and offences of basic intent, with the *dicta* again indicating towards three different interpretations. First, it might be reasoned that the prosecution is only required to prove *actus reus* in crimes of basic intent where the defendant was voluntarily intoxicated; however, such an approach is criticised for effectively transforming crimes of basic intent into strict liability offences wherein no *mens rea* needs be proven.[54] Second, a defendant's voluntary intoxication might itself be regarded as providing the *mens rea* for offences of basic intent on the basis that, through the decision to become intoxicated, the defendant was reckless as to the effects of intoxication on their behaviour and the risks that they might proceed to take in an intoxicated state. However, again, this interpretation is also criticised on a number of points; most notably, it introduces voluntary intoxication as a form of *mens rea* in its own right (as opposed to the *mens rea* of intention or recklessness which is formally required for offences of basic intent). Further, as the intoxication invariably takes place before the subsequent criminal act, this interpretation separates the coincidence of *mens rea* and *actus reus* that is otherwise generally required in order to ascribe legal responsibility for a criminal act.[55]

The third approach to considering the interaction between voluntary intoxication and offences of basic intent is simply to preclude the defendant from relying upon evidence of that intoxication as a relevant consideration when determining whether or not they formed the requisite *mens rea* for the offence. This approach avoids the previous criticisms, either that basic intent offences becomes strict liability offences, that voluntary intoxication itself becomes a form of *mens rea*, or that the coincidence between *mens rea*

[52] *Ibid.,* 474 – 475, 479 & 498.
[53] Arlie Loughnan, *Manifest Madness: Mental Incapacity in the Criminal Law* (Oxford University Press 2012), 187 – 188; citing *Commissioner of Police of the Metropolis v Caldwell* [1982] AC 341, 355.
[54] For example, see Chester N. Mitchell, 'The intoxicated offender – Refuting the legal and medical myths' (1988) 11(1) *International Journal of Law and Psychiatry* 77, 84; Mark T. Thornton, 'Making sense of *Majewski*' (1980-81) 23(4) *Criminal Law Quarterly* 464, 484 – 485.
[55] Alan Dashwood, 'Logic and the Lords in *Majewski*' (1976) *Criminal Law Review* 532; Simon Gardner, 'The importance of *Majewski*' (1994) 14(2) *Oxford Journal of Legal Studies* 279, 281.

and *actus reus* is broken; and, indeed, it is this approach which has been followed in subsequent jurisprudence.[56] The rationale follows that, having *voluntarily* decided to intoxicate themselves to a point where their judgment is impaired and they are liable to act on greater risks, a defendant cannot reasonably appeal to that voluntary decision to absolve themselves of the risks that they later take.[57]

Finally, some relevant points within the extant law can be clarified, before considering how an intoxication defence operates within the context of the hybrid objective / subjective approach to *mens rea*. To begin, a defendant cannot rely on their voluntary intoxication in order to defend against even an offence of specific intent, when they had first formed that particular intention and then subsequently became intoxicated for "courage" in order to carry out that intention.[58] However, there are circumstances of "amoral" intoxication where the defence may still be relied upon; where a defendant is blameless in causing the conditions of their own defence, they are 'no more blameworthy… than is the actor who has made no causal contribution.'[59]

As highlighted above, involuntary intoxication may be pleaded in order to raise reasonable doubt as to whether or not the defendant formed the *mens rea* for offences of either basic or specific intent. However, it must be reiterated that the fact of involuntary intoxication *per se* does not automatically negative *mens rea*, but may be relied upon as evidence that such *mens rea* was absent.[60] Involuntary intoxication is interpreted relatively narrowly such as in circumstances where a person's drink is surreptitiously laced with alcohol or drugs by another,[61] but not merely where the defendant underestimates the effect of some substance that they otherwise voluntarily imbibe.[62]

---

[56] *R v Woods* (1982) 74 Cr App R 312; *R v Richardson and Irwin* (1999) 1 Cr App 392.
[57] Jeremy Horder, 'Sobering up? The Law Commission on criminal intoxication' (1995) 58(4) *Modern Law Review* 534, 540 – 542.
[58] *Attorney-General for Northern Ireland v Gallagher* [1963] AC 349.
[59] Paul H. Robinson, 'Causing the conditions of one's own defense: A study in the limits of theory in criminal law doctrine' (1985) 71(1) *Virginia Law Review* 1, 8.
[60] *R v Kingston* [1995] 2 AC 355, 364 & 377.
[61] *R v Allen* [1988] Crim LR 698.
[62] *R v Eatch* [1980] Crim LR 650.

Further, special mention must be made of "non-dangerous" drugs such as prescription medications, sedatives and soporifics drugs, and intoxication arising from non-compliance with a medicinal regime, for example, when a person fails to take food with medication such as insulin. Voluntary intoxication by alcohol and illegal substances such as cocaine, amphetamines and hallucinogens receives special treatment under the *Majewski* doctrine because of the known deleterious effects that such substances have on people's decision-making and behaviour, and the unreasonableness of a defendant relying on their voluntary assumption of the risks associated with such substances as a means to defend against their own subsequent reckless behaviour. As Morse writes:

> '[T]he culpability in getting drunk – itself not a crime – is the equivalent of actually foreseeing that there might be criminal consequences of one's intoxication. The rationale is that it is common knowledge that intoxication can be disinhibiting or cloud judgment, even if the intoxicated defendant did not actually foresee specifically what those consequences might be.'[63]

The same assumptions cannot (or, at least as a matter of pragmatism, are not) made with regards to "non-dangerous" drugs, in particular medications that are taken under prescription, but also other sedative or soporific drugs such as Valium. In principle, intoxication as a result of these substances, or as a result of a failure to properly comply with a medicinal regime, is treated akin to involuntary intoxication which may be relied upon by the defendant in order to raise reasonable doubt regarding their *mens rea* for offences of either basic or specific intent. Crucially, however, the courts will consider 'whether the defendant was reckless in taking the drugs, and / or failing to eat where medication must be taken before / after / with food.'[64]

Thus, relevant considerations include whether and to what extent a defendant failed to comply with medical instructions, their awareness of risks associated with the substances

[63] Stephen J. Morse, 'Criminal law and addiction' in Pickard H. and Ahmed S. H. (eds.), *The Routledge Handbook of Philosophy and Science of Addiction* (Routledge 2019), 549.

[64] Arlie Loughnan and Nicola Wake, 'Of blurred boundaries and prior fault: Insanity, automatism and intoxication' in Reed A. and Bohlander M. (eds.), *General Defences in Criminal Law: Domestic and Comparative Perspectives* (Ashgate Publishing 2014), 127; citing *R v Burns* (1974) 58 Cr App R 364; *R v Hardie* (1985) Cr App R 157; *R v Bailey* [1983] EWCA Crim 2.

taken, and whether steps could have been taken to prevent their intoxication or mitigate the consequences thereof. Where the defendant is found to have been reckless in taking such "non-dangerous" substances, their intoxication will be treated as voluntary for the purposes of establishing any legal defence, and the approach in *Majewski* is applied.

*

It is submitted that the current approach to intoxication is incorporated into the hybrid objective / subjective approach to *mens rea* relatively simply, having regard to the fact that intoxication does not provide a general defence *per se* but may be relied upon to rebut *mens rea*.

Starting with involuntary intoxication, the fact of a defendant's intoxication would be included amongst the relevant subjective circumstances considered under the second limb of the hybrid test, regardless of whether the offence charged is one of basic or specific intent. This distinction would remain from the current law, such that offences of basic intent are those that can be committed recklessly, whilst offences of specific intent can only be committed with intention. Thus, the relevant question under the second limb of the test would be *whether it is reasonable to expect anybody in the defendant's state of intoxication to appreciate the virtual certainty / unreasonable risk of the relevant offence arising from their actions* as defined under the first limb of the test for whatever offence is charged.

The answer to this question is always likely to be highly context-specific, for example, considering such matters as what substance the defendant used, their level of intoxication and the effects of that substance on their mental faculties, and the nature of the virtual certainty / risk being considered under the first limb of the test. Thus, it will typically remain perfectly reasonable to expect people to appreciate the gravest, most obvious and / or harmful certainties / risks notwithstanding a high level of intoxication, whereas it may be more reasonable to expect that people would not appreciate other less serious, obvious or harmful risks as they become increasingly intoxicated.

With regards to voluntary intoxication, a defendant would be precluded from raising their intoxication amongst the relevant subjective circumstances under the second limb of the test whenever the offence charged is one of basic intent, just as they are equally unable to rely upon intoxication to refute the *mens rea* of recklessness under the current law. However, if the offence charged is one of specific intent, then the voluntarily intoxicated defendant would remain permitted to raise their intoxication under the second limb of the test, precisely as described above in relation to involuntary intoxication. The crucial difference follows that, whereas a successful plea of involuntary intoxication will provide a complete defence to offences of both basic and specific intent, a successful plea of voluntary intoxication only provides a partial defence, reducing a charge of specific intent to its lesser, basic intent equivalent – *e.g.,* reducing a charge of murder to manslaughter.

An additional step in the above formulation is included when considering voluntary intoxication by "non-dangerous" substances. Here, it must first be considered whether or not the defendant has been reckless in consuming and becoming intoxicated by the relevant substances, applying the hybrid formulation of recklessness. Thus, the first limb of the test asks whether there was an objectively unreasonable risk that the defendant would become intoxicated by consuming the relevant substance in the manner and quantity taken.

Finally, the second limb of the test asks *whether it is reasonable to expect anybody in the defendant's (sober) circumstances to appreciate that unreasonable risk associated with the substance that they took.* If both limbs of the hybrid test are answered in the affirmative, then the defendant's intoxication is treated as voluntary and may only be considered amongst their relevant subjective circumstances in relation to an offence of specific intent, as described above. If either limb of the test is answered in the negative, then the defendant's intoxication is treated as involuntary and may be considered amongst their relevant subjective circumstances for the purposes of any offence.

### *11.3.4. Insanity*

Insanity is the quintessential example of a defence that potentially impacts upon all three of the capacities underlying *mens rea* – they are, the capacities for reasons responsiveness, for ordinary self-control, and for appreciating the nature of one's conduct. The defence of insanity is a predominantly common law defence 'governed by the so-called M'Naghten rules' developed in 1843.[65] The defendant in *Daniel M'Naghten's Case*[66] killed the secretary to the Prime Minister, mistakenly believing it to be the Prime Minister himself, because he was suffering from the insane delusion that he was being persecuted by the Tory party with attempts being made upon his life. With public outcry at the defendant's acquittal at trial, the legislative House of Lords put questions to its judicial counterparts in order to elucidate the defence of insanity, and the judicial House of Lords responded with the "M'Naghten rules":

> '[T]o establish a defence on the ground of insanity, it must be clearly proved that, at the time of the committing of the act, the party accused was labouring under such a defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing; or, if he did know it, that he did not know he was doing what was wrong.'[67]

First, the defendant must suffer from a defect of reason at the time of the offence, which is 'more than a momentary confusion of absent-mindedness; a deprivation of reasoning power is required.'[68] In *R v Clarke,*[69] for example, the defendant was in a depressed state induced by low blood sugar and her diabetic condition, and mindlessly and inadvertently placed an item from a supermarket shelf into her shopping bag. She was subsequently prosecuted for theft and pleaded an absence of *mens rea,* whilst the trial judge directed that she was raising a plea of insanity. However, the Court of Appeal determined that the defendant's sanity had not been put into issue; '[s]he was not asserting a defect of reason arising from a disease of the mind. Rather, because of absent-mindedness associated with

[65] Steven Yannoulidis, *Mental State Defences in Criminal Law* (Routledge 2016), 9.
[66] *Daniel M'Naghten's Case* (1843) 8 ER 718.
[67] *Ibid*., 722.
[68] Richard Card and Jill Molloy, *Card, Cross & Jones Criminal Law* (22nd ed. Oxford University Press 2016), 614.
[69] *Clarke* [1972].

depression, she failed to exercise a faculty she still possessed.'[70] The relevant issue was therefore a matter of the existence or otherwise of *mens rea* and not any defect of reasoning.

Second, that defect of reasoning must arise as a result of a "disease of the mind"; this is a decidedly legal term and not one which follows any medical definition or approach.[71] Indeed, as is discussed further below in this section, the law has taken an appreciably unprincipled approach to determining precisely what is considered to be a disease of the mind. In general terms, a disease of the mind may be 'organic or functional, permanent or transitory or intermittent,'[72] provided that it is in operation at the time of the alleged offence and impacts upon the individual's 'mental faculties of memory, reason and understanding.'[73] The term is generally understood to include most serious mental disorders albeit,[74] somewhat paradoxically, has also been deemed to include such physical disorders as hyperglycaemia, sleepwalking and epilepsy on account of their impact upon normal mental functioning.[75]

Third, the defect of reason arising from a disease of the mind must be such that the defendant does not know the nature and quality of their act. This requirement has generally been interpreted strictly to refer to knowledge of the 'physical nature and quality of the act and not to its moral or legal quality.'[76] The requirement has not received particular attention in jurisprudence, and appears to mean simply that the defendant 'did not know what he was doing.'[77] Popular illustrations provided in academia include the hypothetical scenarios where one person kills another under the delusion that he is breaking into a jar,[78] cuts another's head off under the delusion that they were cutting a

[70] Andrew P. Simester, John R. Spencer, Findlay Stark, G. R. Sullivan and Graham J. Virgo, *Simester and Sullivan's Criminal Law: Theory and Doctrine* (7th ed. Hart Publishing 2019), 767.
[71] David Ormerod and Karl Laird, *Smith, Hogan, and Ormerod's Criminal Law* (15th ed. Oxford University Press 2018), 292.
[72] Loughnan (2012), 118.
[73] *R v Kemp* [1957] 1 QB 399, 407; see also *R v Hennessy* [1989] 1 WLR 287, 292.
[74] *Bratty v Attorney-General for Northern Ireland* [1963] AC 386, 412.
[75] Loughnan (2012), 118; citing *R v Burgess* [1991] 2 QB 92.
[76] Ormerod and Laird (2018), 298.
[77] *R v Sullivan* [1984] 1 AC 156, 173.
[78] James Fitzjames Stephen, *A Digest of the Criminal Law* (8th ed. Sweet and Maxwell 1947), 6.

loaf of bread,[79] or because it would be 'great fun to see him looking for it when he woke up.' [80] As Allen and Edwards explain, 'such a person does not understand the consequences of his acts.'[81]

The fourth element of the insanity defence operates further or in the alternative to the third, and requires that the defect of reason arising from a disease of the mind is such that the defendant did not know that their actions were wrong. Here, again, wrongfulness has been given a relatively narrow interpretation to refer to the fact that the particular actions in question were illegal, *i.e.,* amounted to a criminal offence.[82] The rationale follows that courts are only competent to determine whether a defendant's actions are in accordance with law and not morality, whereas leaving the question of morality to the jury would introduce the inevitable query of which or whose morality was being imposed.[83] This restrictive interpretation of wrongfulness has also come under significant academic criticism, however, some of which is considered further in this section, below. It is notable that the High Court of Australia has declined to follow this rule, expanding wrongfulness to include that which is wrong 'according to the ordinary standards of reasonable people.'[84]

*

The manner in which the defence of insanity fits within the hybrid approach to *mens rea* defended in this thesis is evident. Where it has been argued that *mens rea* is underpinned by three capacities – for reasons responsiveness, ordinary self-control, and appreciating the nature of one's conduct – the insanity defence essentially operates where, arising from some "disease of the mind", one or more of these three capacities were undermined or overwhelmed at the time of the alleged offence. The very first condition of the *M'Naghten* rules requires that the defendant is suffering from a defect of reason; in his seminal work on the subject, Fingarette defines legal insanity as existing when 'the individual's mental

[79] J. W. Cecil Turner, *Kenny's Outlines of Criminal Law* (19th ed. Cambridge University Press 1966), 76.
[80] James Fitzjames Stephen, *A History of the Criminal Law of England – Vol II* (Macmillan & Co. 1883), 166.
[81] Michael Allen and Ian Edwards, *Criminal Law* (15th ed. Oxford University Press 2019), 163 – 164.
[82] Ormerod and Laird (2018), 299 – 300.
[83] See *R v Windle* [1952] 2 QB 826, 833 – 834; *R v Johnson* [2007] EWCA Crim 1978.
[84] *Ibid*., 300; citing *Stapleton v R* (1953) 86 CLR 358; *R v Weise* [1969] VR 953, 960.

makeup at the time of the offending act was such that, with respect to the criminality of his conduct, he substantially lacked capacity to act rationally.'[85] In this respect, it is appreciable how the insanity defence may arise when, due to some disease of the mind, the defendant's capacity for thinking rationally and being responsive to reason is not functioning properly.

The third limb of the *M'Naghten* rules concerning the defendant's lack of knowledge as to the nature and quality of their act can be related to the capacity for ordinary self-control. In *Loake v Director of Public Prosecutions*[86] the defendant was charged with harassment on account of a large number of messages that she had sent to her estranged husband. However, the defendant suffered from dementia such that she could not remember each time that she attempted to make contact to her husband, as a result of which the High Court determined that the defendant 'would not know the nature and quality of their act.'[87] This link is clearer when considering the defence of insane automatism. Discussed below in section 11.3.5, automatism essentially comprises of the defendant's inability to control their physical movements, whether due to being unconscious or the result of some other impairment over the control of bodily actions, such as spasm or reflex actions.[88] Crucially, where such a loss of voluntary control arises from an "internal" cause, the defence is categorised as insane automatism and receives similar treatment to the defence of insanity.[89] For example, the defendant in *R v Sullivan*[90] was physically violent as a result of a psychomotor epileptic seizure for which he later had no recollection, for which the House of Lords determined that the appropriate verdict was one of not guilty by reason of insanity. In particular, the defendant's condition resulted in his not knowing the nature or quality of his violent actions which, consequently, were not under his ordinary self-control.[91]

[85] Herbert Fingarette, *The Meaning of Criminal Insanity* (University of California Press 1972), 211.
[86] *Loake v Director of Public Prosecutions* [2018] QB 998.
[87] *Ibid*., 1012.
[88] Allen and Edwards (2019), 167 – 168.
[89] See *Bratty* [1963], 409 – 410; *R v Quick* [1973] QB 910, 922.
[90] *Sullivan* [1984].
[91] See further Ronnie D. Mackay, '"Nature", "quality" and *mens rea* – Some observations on "defect of reason" and the first limb of the M'Naghten rules' (2020) 7 *Criminal Law Review* 588.

Finally, the fourth element of the *M'Naghten* rules concerning the defendant's lack of knowledge about the wrongfulness of his actions may be related to the capacity to appreciate the nature of one's conduct as it relates to the offence charged. In particular, the English approach that such wrongfulness relates only to the *legal* wrongfulness of a given act is broadly supported by the hybrid approach to *mens rea*. The second limb of the hybrid test is particularly interested in the reasonableness of expecting anybody sharing the defendant's circumstances to appreciate the nature of their actions *as they relate to the offence in question*. Thus, with this element making direct reference to the offence for which a defendant is being tried, and with subjective circumstances and characteristics being relevant only insofar as they concern the three capacities for responsibility, (in particular, the capacity to appreciate the nature and consequences of conduct relating to that offence), it is defensible that the wrongfulness limb of the *M'Naghten* rules is also restricted to consider the legal (as opposed to moral) wrongfulness of any conduct.

Thus, by reference to the three capacities underlying *mens rea* generally, it may readily be appreciated how and why the defence of insanity operates based upon the fact that one or more of these crucial capacities has been overwhelmed or incapacitated on account of a disease of the mind. However, it ought to be noted that, in keeping with the broader rejection of pure subjectivity supported throughout this thesis, the requirements within the *M'Naghten* rules of a defendant's lack of knowledge as to either the nature and quality or wrongfulness of their actions must be determined according to the hybrid approach to knowledge described in section 10.3 of this thesis, above. The objective nature of the required knowledge is already determined by the *M'Naghten* rules – *i.e.,* knowledge as to either the quality and nature or wrongfulness of an action. Therefore, the pertinent question is will always be the second limb of the hybrid test, asking whether or not it is reasonable to expect anybody in the defendant's circumstances (including sharing their defect of reasoning from a disease of the mind) to appreciate the quality and nature of their actions or the certainty that they were wrong.

*

Although the following section 12.3.1 of this thesis considers reform of the verdict of not guilty by reason of insanity, it is pertinent to note some of the wider reforms to the defence of insanity that have been recommended by the Law Commission[92] and / or adopted in other common law jurisdictions. These recommendations would not only be entirely supported by the present thesis, but encapsulate the manner in which the insanity defence is considered to apply by reference to the three capacities underlying *mens rea*. One of the key criticisms against the defence of insanity as currently formulated concerns the unprincipled manner in which different medical conditions have and have not been included within the defence. On the one hand, the defence appears to offer protection to those suffering from only the most serious mental disorders, including severe psychotic, delusional or dissociative conditions. Conversely, neurotic, emotional and volitional disorders such as depression, anxiety and obsessive / compulsive disorders are excluded. On the other hand, however, 'the defence is also surprising because… it does, strangely, include defendants with common physical illnesses such as diabetes, sleepwalking and epilepsy.'[93] The defence can therefore appear, at once, both over- and under-inclusive.[94]

Moreover, continuing from the previous discussion, the *M'Naghten* rules fall under considerable criticism both regarding the nature of knowledge (*i.e.,* actual or the capacity for) and the interpretation of "wrongfulness". Regarding the former, it is arguable that the requirement that the defendant lacks *actual* knowledge of the nature and quality or wrongfulness of their actions is unduly restrictive and under-inclusive because it places the principal focus upon the existence of a particular state of mind appearing within the mind of a person whom, by the first two limbs of the *M'Naghten* rules, has already been determined to be suffering from a defect of reason arising from a disease of the mind. Consequently, it is popularly argued that the focus should instead be upon whether the defendant retained the *capacity* to appreciate the nature and quality or wrongfulness of their actions.[95] Such a move would be broadly supported by the present thesis under

[92] Law Commission, *Criminal Liability: Insanity and Automatism – A Discussion Paper* (Law Commission 2013).
[93] Janet Loveless, Mischa Allen and Caroline Derry, *Complete Criminal Law: Text, Cases, and Materials* (7th ed. Oxford University Press 2020), 326.
[94] See further Mackay and Reuber (2007); Mackay and Mitchell (2006); Arlie Loughnan, '"Manifest madness": Towards a new understanding of the insanity defence' (2007) 70(3) *Modern Law Review* 379.
[95] See Victor Tadros, *Criminal Responsibility* (Oxford University Press 2007), Ch. 12.

which mental capacities (and not the existence of specific states of mind) are of critical relevance to criminal responsibility.

Concerning the interpretation of wrongfulness, the key criticism is encapsulated in the report of the Butler Committee on Mentally Abnormal Offenders[96] which stated that:

> '[K]nowledge of the law is hardly an appropriate test on which to base ascription of responsibility to the mentally disordered. It is a very narrow ground of exemption since even persons who are grossly disturbed generally know that murder and arson are crimes.'[97]

It is further noted that ignorance of the law provides no defence for neurotypical adults who are not suffering from any defect of reason. It is therefore somewhat unprincipled that ignorance of the fact that an individual's actions are legally wrong does become a defence if such ignorance arises from some disease of the mind, but not where such ignorance simply reflects a lack of knowledge possessed by an ordinary or neurotypical individual. Moreover, empirical evidence overwhelmingly suggests that psychiatrists appearing as expert witnesses in trials involving the insanity defence are more typically approaching the issue of wrongfulness as a moral rather than legal question.[98]

Notably, the courts in certain States of Australia have departed from restraining both the requirement of knowledge and the meaning of wrongfulness as is the case in the UK. Concerning the nature of the defendant's knowledge, the case of *R v Porter*[99] provides:

> '[T]he question is *whether he was able to appreciate* the wrongness of the particular act he was doing at the particular time… if through a disease or

[96] Butler Committee, *Report of the Committee on Mentally Abnormal Offenders* (Cmnd 6244, 1975).
[97] *Ibid*., [18.8]; see also Ronnie D. Mackay, *Mental Condition Defences in the Criminal Law* (Clarendon Press 1995), 97.
[98] See Gerry Kearns and Ronnie D. Mackay, 'More fact(s) about the insanity defence' (1999) *Criminal Law Review* 714, 723.
[99] *R v Porter* [1936] 55 CLR 182.

> disorder of the mind *he could not think rationally* of the reasons which to ordinary people make that act right or wrong?
>
> 'If through the disordered condition of the mind *he could not reason about the matter with a moderate degree of sense and composure* it may be said that he could not know that what he was doing was wrong.'[100]

Whereas the *M'Naghten* rules are interpreted in the UK and some Australian States as requiring actual knowledge, other States in Australia (including Queensland and Tasmania) follow the capacity-based approach expounded in *Porter*.[101] Furthermore, on the interpretation of wrongfulness, it has already been highlighted above where the Australian courts have departed from the English approach in the case of *Stapleton v R*, considering whether or not a defendant was able to '"think rationally" about whether their act was *morally* wrong.'[102] Again, further states including Queensland, Tasmania and Western Australia have developed this moral wrongfulness into a capacity test requiring that the defendant was 'deprived of the *capacity* to know that [they] ought not to do the act.'[103] Notwithstanding the fact that the current English approach to the *M'Naghten* rules is perfectly appreciable within the present thesis by reference to the three capacities underlying *mens rea*, it is nonetheless submitted that the relevant Australian developments of the law would be more supportable still, not least for shifting the focus of the insanity defence away from subjective knowledge and towards mental capacities.[104]

Similarly advocating for a capacity-based approach, the Law Commission of England and Wales proposes a new defence to replace insanity where the defendant is not responsible by reason of recognised medical condition. Specifically, the defendant would be required

---

[100] *Ibid*., 189 – 190.
[101] See further Yannoulidis (2016), 16 – 18.
[102] *Ibid*., 15.
[103] *Ibid*; for a fuller comparative approach, see Keith J. B. Rix, 'Prizing open the door to justice: Reform of the "wrongfulness limb" of the *M'Naghten* Rules' in Livings B., Reed A. and Wake N. (eds.), *Mental Condition Defences and the Criminal Justice System: Perspectives from Law and Medicine* (Cambridge Scholars Publishing 2015).
[104] See also Mark Hathaway, 'The moral significance of the insanity defence' (2009) 73(4) *Journal of Criminal Law* 310, 316 – 317.

to adduce expert evidence that, at the time of the alleged offence and as a result of a qualifying recognised medical condition, they 'wholly lacked the capacity:

(i) rationally to form a judgment about the relevant conduct or circumstances;
(ii) to understand the wrongfulness of what he or she is charged with having done; or
(iii) to control his or her physical acts in relation to the relevant conduct or circumstances.'[105]

The parallels between this proposal and how the current insanity defence has been reasoned in this thesis to operate within the hybrid approach to *mens rea* are plain and clear to appreciate. In particular, the first and third of the capacities identified by the law commission respectively match the capacities for being responsive to reasons and for ordinary self-control that are rebuttably presumed to exist for all adults within the concept of volition. Although not an exact correlate, the second capacity identified by the law commission to understand the wrongfulness of one's actions broadly relates to the capacity to appreciate the nature and consequences of one's conduct that is of central importance under the second limb of the hybrid approach to *mens rea*. Thus, the Law Commission proposes reforms to the insanity defence that are not merely supported by this present thesis, but capture the very essence of what is at issue under the revised hybrid approach to *mens rea – i.e.,* mental capacities as opposed to subjective states of mind.

Moreover, the approach proposed by the Law Commission and supported by this thesis would arguably address many of the issues that have been identified within the existing defence of insanity. Regarding the two examples mentioned above, the defence of non-responsibility by reason of recognised medical condition is notably more inclusive than the current defence of insanity whilst, being tied to recognised medical conditions, clearly delineates those conditions that may be included within the defence. The Law Commission specifies that any such condition 'must be one which *could* cause the

[105] Law Commission (2013), 91; for extensive discussion, see Jesse Elvin and Claire de Than, 'The boundaries of the insanity defence: The legal approach where the defendant did not "know that what he was doing was wrong"' in Livings B., Reed A. and Wake N. (eds.), *Mental Condition Defences and the Criminal Justice System: Perspectives from Law and Medicine* (Cambridge Scholars Publishing 2015).

individual accused the lack of capacity which he or she claims in the particular case.'[106] The proposed reform would therefore make no distinction between physical and mental conditions, and recognised disorders including epilepsy, Alzheimer's disease, schizophrenia, bipolar disorder and clinical depression would all be potentially relevant. Nevertheless, '"abnormal" physical and mental states which do not amount to a recognised medical condition would not fall within the defence.'[107] Thus, whereas post-traumatic stress disorder arising from bereavement could be relevant, ordinary grief would not; similarly, Alzheimer's disease could be relevant to a defendant's violent reaction, but simply being "hot-headed" or quick to temper would not.

The Law Commission proposal also reforms what is meant by "wrongfulness" under the second relevant capacity of the recognised medical condition defence. Following popular academic criticism, the Law Commission agrees that confining the meaning of this word to purely legal wrongs is unduly narrow; however, widening the concept to be interpreted as referring to moral wrongs 'begs the question whose morality is to be used as the standard by which the accused's appreciation is judged.'[108] The approach taken in Canadian law was therefore approved, which requires that the defendant 'need only appreciate that the act was something he or she ought not to do.'[109] It is submitted that this brings the concept of wrongfulness considerably closer that the capacity that is at issue under the second limb of the hybrid test – the capacity to appreciate the nature of one's actions as it relates to the specific *mens rea* of the offence charged. The argument follows, that which a particular defendant "ought not to do" *is* the prohibited *actus reus* of a criminal offence in conjunction with the requisite *mens rea*. Therefore, when the second limb of the hybrid approach to *mens rea* concludes that it is indeed reasonable to expect anybody in the defendant's position to appreciate the nature of their conduct, the particular nature being referred to is the manner in which that conduct relates to the *mens rea* of the offence charged, *i.e.,* a component of that which the defendant *ought not to do*.

[106] *Ibid*., 65.
[107] *Ibid*., 66.
[108] *Ibid*., 56.
[109] *Ibid*; *R v Codère* (1916) 12 Cr App R 21; *R v Chaulk* [1990] 3 SCR 1303.

Finally, the Law Commission's proposals are considerably preferable to the current insanity defence for the practical reason that it removes the requirement to prove that a defendant subjectively knew that their conduct was wrong. Under the hybrid approach here discussed, "knowledge" within the current insanity defence would need to be addressed following the hybrid approach to knowledge described in section 10.3 of the thesis, above. However, it might fairly be argued as being unnecessarily complicated for a jury to bring into the question of a defendant's knowledge, for the purposes of the insanity defence, the test of whether or not it is reasonable to expect anybody in the same circumstances to appreciate the nature of their actions. This results in a degree of circularity as, if the defendant is indeed suffering under a defect of reason arising from a disease of the mind, it is increasingly likely to be *un*reasonable to expect anybody in the same circumstances to appreciate that their conduct was wrongful. By focussing instead upon whether a recognised medical condition extinguishes one or more of three relevant mental capacities, the Law Commission's proposal is not only more congruent with the broader approach argued within this thesis, but even resolves any alleged circularity in the application of the hybrid approach to knowledge within the defence of insanity.

### *11.3.5. Automatism*

The defence of automatism is available when 'movements or actions of the defendant at the material time were wholly involuntary… [with the] complete destruction of voluntary control.'[110] The defence comes down to the *total* inability to control the actions that result in criminal conduct. Crucially, as Simester, Spencer, Stark, Sullivan and Virgo confirm:

> '[U]nconsciousness or impaired consciousness is not *required* for the defendant to be absolved of responsibility… [w]hether she was conscious or unconscious, what is essential to the denial of responsibility for a defendant's involuntary behaviour is that *she was unable deliberatively to control that behaviour and to prevent it from occurring*.'[111]

[110] *R v Coley, McGhee and Harris* [2013] EWCA Crim 223, [22].
[111] Simester, Spencer, Stark, Sullivan and Virgo (2019), 122 (original emphasis).

Consequently, whilst the degree of control that a person is able to exercise over their actions for the defence of automatism must be entirely abrogated, and not merely 'impaired, reduced, or partial',[112] there is no necessary link between that loss of control and (un)consciousness.[113] Indeed, in *Bratty v Attorney-General for Northern Ireland* the House of Lords subsumed within automatism both actions that are wholly involuntary *or* performed unconsciously. The Court described the defence as covering both:

> '[A]n act which is done by the muscles without any control by the mind such as a spasm, a reflex action or a convulsion; or an act done by a person who is not conscious of what he is doing such as an act done whilst suffering from concussion or whilst sleepwalking.'[114]

The distinction between involuntary and unconscious automatism may be understood more clearly by reference to how the defence operates in either case. Loughnan explains that automatism consists of a claim of incapacity with both a mental and physical dimension; 'the defendant is claiming that the mental element of the *actus reus* (voluntariness) is lacking or, alternatively, that the physical element of the *mens rea* (consciousness) is lacking.'[115] In either case, however, the "essence" of the automatism defence lies in the claim that the defendant was unable to control their bodily movements which amounted to the *actus reus* of the offence charged.[116]

A distinction within the defence of automatism that is eminently more important, however, concerns whether or not the defendant's involuntariness arose as a result of an internal or external cause. Where automatism arises from some internal cause – *i.e.,* a "disease of the mind" within the meaning of the insanity defence – then the defence is treated as insane automatism and the special verdict of not guilty by reason of insanity is appropriate. Examples of insane automatism in jurisprudence include involuntariness arising from

---

[112] *Attorney-General's Reference (No. 2 of 1992)* [1994] QB 91, 105.
[113] *Watmore v Jenkins* [1962] 2 QB 572, 586.
[114] *Bratty* [1963], 409; see further John Rumbold, *Automatism as a Defence in Criminal Law* (Routledge 2018), 81 – 83.
[115] Loughnan (2012), 127; citing Norval Morris, 'Somnambulistic homicide' (1951) 5(1) *Res Judicatae* 29; see also Rumbold (2018), 82.
[116] Andrew Ashworth and Jeremy Horder, *Principles of Criminal Law* (7th ed. Oxford University Press 2013), 89.

psychomotor epilepsy,[117] hyperglycaemic states resulting from a failure to take sufficient insulin,[118] sleepwalking,[119] and arteriosclerosis.[120] Conversely, where the cause of a defendant's involuntary action was something external, the defence of sane automatism is available entitling the defendant to a full acquittal. Quintessential hypothetical examples of sane automatism are given in *Hill v Baxter*[121] as being where a driver reacts involuntarily because they were 'struck by a stone or overcome by a sudden illness; or the car was temporarily out of control by his being attacked by a swarm of bees.'[122]

The internal / external cause dichotomy has been far from simple to apply, has led to some particularly counterintuitive results, and is roundly criticised for being unhelpful "lawyer speak".[123] For example, whereas epileptic seizures and parasomnias have clearly internal causes, they can also be 'precipitated by external factors.'[124] Furthermore, the infamously peculiar situation has emerged whereby involuntary behaviour caused by hypoglycaemia resulting from an insulin overdose or lack of food is treated as an external cause, attracting the defence of sane automatism with a full acquittal. Meanwhile, involuntary behaviour caused by hyperglycaemia resulting from a lack of insulin or an insulin-secreting tumour is regarded as being an internal cause, attracting the defence of insane automatism and the special verdict of not guilty by reason of insanity.[125] Finally, it is important to note that the defence of automatism is precluded where the defendant is responsible for being in the state of an automaton, for example, where they are voluntarily intoxicated.[126]

---

[117] *Bratty* [1963]; *Sullivan* [1984]; see further G. M. Paul and K. W. Lange, 'Epilepsy and criminal law' (1992) 32(2) *Medicine, Science and the Law* 160.
[118] *Hennessy* [1989]; *R v Bingham* [1991] Crim LR 43.
[119] *Burgess* [1991].
[120] *Kemp* [1957].
[121] *Hill v Baxter* [1958] 1 QB 277.
[122] *Ibid*., 282 – 283; see further Law Commission (2013), Ch. 5.
[123] For example, see Irshaad Ebrahim, Peter Fenwick, Richard Marks and Kevin W. Peacock, 'Violence, sleepwalking and the criminal law: Part 1: The medical aspects' (2005) (Aug) *Criminal Law Review* 601, 602 – 603.
[124] Mark A. Turner and Nicholas F. Moran, 'Automatism: The ictus, the character, and the law' in Livings B., Reed A. and Wake N. (eds.), *Mental Condition Defences and the Criminal Justice System: Perspectives from Law and Medicine* (Cambridge Scholars Publishing 2015), 218; citing Mackay (1995), 38; Irshaad Ebrahim and Peter Fenwick, 'Sleep related automatism and the law' (2008) 48(2) *Medicine, Science and the Law* 124, 127.
[125] Loughnan (2012), 129; Alex Samuels, 'The diabetic driver' (2019) 59(4) *Medicine, Science and the Law* 282; Vincent Marks, 'Hypoglycaemia and automatism' (2015) 55(3) *Medicine, Science and the Law* 186.
[126] *Ibid*, 131 – 133; citing *Majewski* [1977], 487; *Hardie* (1985), 162.

*

As with the defence of insanity, it is not difficult to appreciate how the defences of both insane and sane automatism fit within the broader theory of responsibility defended in this thesis. Specifically, the automatism defence claims that the defendant was entirely unable to control their actions, whether by reason of unconsciousness or some other internal or external factor that abrogated the voluntariness of their actions. Therefore, this defence unequivocally relates to the capacity for ordinary self-control that is presumed to exist for all adults unless that presumption is rebutted; the defence of automatism is the very essence of the claim that a defendant lacked the capacity for self-control when committing the *actus reus* of a given offence.

In addition, the recommendations for reform to the defence of automatism from the Law Commission are yet more supportable under the present thesis. Specifically with regards to automatism, any loss of control arising from a recognised medical condition would fall under the newly recommended defence of not responsible by reason of recognised medical condition. Subsequently, any such total loss of control arising due to anything other than a recognised medical condition would fall under the defence of automatism. In practice, the Law Commission explains, the automatism defence is 'likely to be applicable in relation to automatic reflex reactions, or to transient states or circumstances; if a person's condition persists and worsens it might then qualify as a recognised medical condition.'[127] Such a move would resolve many of the current criticisms concerning the somewhat unprincipled separation of sane and insane automatism in the extant law.[128] Meanwhile, under the Law Commission's recommendations, both loss of control due to a recognised medical condition and the remaining defence of non-medical automatism each relate to the defendant's capacity for ordinary self-control yet more clearly than the current law.

[127] Law Commission (2013), 122.

[128] Ronnie D. Mackay, 'An anatomy of automatism' (2015) 55(3) *Medicine, Science and the Law* 150, 154; Lisa Claydon, 'Reforming automatism and insanity: Neuroscience and claims of lack of capacity for control' (2015) 55(3) *Medicine, Science and the Law* 162.

### *11.3.6. Diminished Responsibility and Loss of Control*

The defences of diminished responsibility[129] and loss of control[130] are both creations of statute. The defences are only available with respect to the charge of murder, and do not provide an acquittal but operate to reduce the defendant's conviction to the lesser charge of manslaughter.[131] In this regard, the application of these defences does not suggest that the defendant was not at all responsible for committing the offence charged, but that they were not *wholly* responsible. Both of these defences are of considerable practical importance as a conviction for murder carries a mandatory sentence of life imprisonment, whilst the judge retains discretion in sentencing for the lesser offence of manslaughter.[132]

The requirements for the defence of diminished responsibility are set out in section 2 of the Homicide Act 1957, as amended. The defendant must be suffering from an 'abnormality of mental functioning' arising from a recognised medical condition, which 'provides an explanation' for the defendant's conduct in killing another, and which 'substantially impairs' the defendant's ability either to understand the nature of their conduct, to form a rational judgment or to exercise self-control. The term "abnormality of mental functioning" has not received extensive judicial attention, but is generally given a considerably wider interpretation than "disease of the mind" within the defence of insanity.[133] Lord Parker CJ stated in the leading case of *R v Byrne*:[134]

> '"Abnormality of the mind", which has to be contrasted with the time-honoured expression in the M'Naghten Rules, "defect of reason", means a state of mind so different from that of ordinary human beings that the reasonable man would term it abnormal. It appears to us to be wide enough to cover the mind's activities in all its aspects, not only the perception of physical acts and matters and the ability to form a rational judgment

[129] Homicide Act 1957, s. 2.
[130] Coroners and Justice Act 2009, s. 54.
[131] Herring (2020), 235 – 236 & 254.
[132] *Ibid*.
[133] See Ronnie D. Mackay, 'The abnormality of mind factor in diminished responsibility' (1999) (Feb) *Criminal Law Review* 117.
[134] *R v Byrne* [1960] 2 QB 396.

whether an act is right or wrong, but also the ability to exercise will-power to control physical acts in accordance with that rational judgment.'[135]

The requirement for "substantial impairment" is undoubtedly a lower standard than the total impairment that is required for the insanity defence; however, the courts have been somewhat inconsistent regarding what amounts to "substantial".[136] The dominant view previously maintained that an impairment was substantial if it was 'more than minimal or trivial.'[137] However, a contrasting view considers substantial to mean that the impairment must be held to some higher standard, albeit still less than a total impairment.[138] Most recently in *R v Golds*,[139] the Supreme Court affirmed the latter interpretation in considering that 'whilst the impairment must indeed pass the merely trivial before it need be considered, it is not the law that *any* impairment beyond the trivial will suffice.'[140]

Finally, the legislation on diminished responsibility requires that the defendant is substantially impaired with regards to their ability to understand the nature of their conduct, to form a rational judgment, or to exercise self-control. The parallels may readily be drawn between these "abilities" and the three "capacities" in issue under the reforms to the insanity defence recommended by the Law Commission. Insofar as the defence of diminished responsibility is more inclusive than insanity with regards to its requisite components but, at the same time, results in a more stringent consequence – *i.e.,* conviction for manslaughter as opposed to not guilty by reason of insanity – the defence might fairly be regarded as providing a limited buffer zone in circumstances where the defence of insanity is not sufficiently made out.

The defence of loss of control is governed under section 54 of the Coroners and Justice Act 2009. The defence requires that the defendant's actions in killing another resulted

---

[135] *Ibid*., 403.
[136] See Barry Mitchell and Ronnie D. Mackay, 'The gold standard of substantial impairment' (2015) 4 *Archbold Review* 7.
[137] *Ibid*., 7; citing *R v Lloyd* [1967] 1 QB 175; see also *R v Brown* [2011] EWCA Crim 2796.
[138] *Ibid*; citing *R v Simcox* [1964] Crim LR 402.
[139] *R v Golds* [2016] UKSC 61.
[140] *Ibid*., [43]; see further Matthew Gibson, 'Diminished responsibility in *Golds* and beyond: Insights and implications' (2017) 7 *Criminal Law Review* 543; Karl Laird, 'Homicide: *R v Golds (Mark Richard)* Supreme Court' (2017) 4 *Criminal Law Review* 316.

from their 'loss of self-control', that loss of self-control had a qualifying trigger, and that any other person of the defendant's age, sex, in the same circumstances and 'with a normal degree of tolerance and self-restraint… might have reacted in the same or in a similar way.'[141] Section 55 of the Act describes the relevant qualifying triggers, consisting of the defendant being in fear of serious violence towards themselves or another from the victim, the victim having said or done things which both 'constituted circumstances of an extremely grave character' and caused the defendant to have a 'justifiable sense of being seriously wronged,' or a combination of the two triggers.

Beginning with the element of loss of control itself, section 55(2) of the 2009 Act provides that it does not matter whether the loss of control was sudden or arose over time, in stark contrast to the replaced defence of provocation.[142] However, section 55(4) qualifies this point by maintaining that the defendant must not have 'acted in a considered desire for revenge.' These changes are intended to redress criticisms under the previous law of provocation which, in requiring loss of control to be sudden and temporary, advantaged male defendants who acted in rage but provided no defence to the (more often female) victims of abusive relationships whose self-control more typically eroded over time.[143]

The courts have been reticent to dictate whether or not loss of control must be total or partial;[144] however, it is important to recall that the defence operates against the charge of murder and, therefore, presupposes that the defendant must have formed the requisite *mens rea* for that offence. Consequently, Herring argues that it cannot be required that the defendant has either '*completely* lost control of her actions or was so angry that she was not aware of what she was doing, because if either of these were true, then the defendant would not have the *mens rea* or *actus reus* of murder, in which case there would be no need to have the defence.'[145]

[141] Coroners and Justice Act 2009, s. 54(1).
[142] See also *R v Dawes* [2013] EWCA Crim 322, [54].
[143] Barry Mitchell, 'Years of provocation, followed by a loss of control' in Zedner L. and Roberts J. V. (eds.), *Principles and Values in Criminal Law and Criminal Justice: Essays in Honour of Andrew Ashworth* (Oxford University Press 2012), 125 – 127.
[144] For example, see *R v Gurpinar* [2015] EWCA Crim 178, [20].
[145] Herring (2020), 237 – 238.

Concerning the second element, it is necessary to prove that the loss of control was caused by a qualifying trigger;[146] where this element is not proven, the defence cannot be left to the jury.[147] In addition to setting out the qualifying triggers, section 55 of the 2009 Act further specifies three rules to be applied in determining whether the defendant's loss of control is attributable to a qualifying trigger. First, any fear of serious violence must be disregarded insofar as it was caused by something the defendant did to provide an excuse for such violence. Second, any sense of being seriously wronged is similarly disregarded where the defendant incited that thing to be said or done for the purpose of providing an excuse for violence. And third, 'the fact that a thing done or said constituted sexual infidelity is to be disregarded.'[148] This latter stipulation was included for important reasons of policy, although has equally received criticism for being somewhat arbitrary in its exclusion and unrealistic regarding the emotions that surround infidelity and relationship breakdown.[149] Specifically, this inclusion exists to preclude overwhelmingly male defendants from pleading a sudden loss of control in response to their partner's infidelity, which previously caused considerable injustice for the victims of such violent attacks under the replaced defence of provocation.[150]

The third element of the loss of control defence requires that any other person sharing the defendant's sex and age, in the same circumstances as the defendant and possessing a normal degree of tolerance and self-restraint might have reacted in a similar manner. Section 54(3) of the 2009 Act further specifies that the relevant circumstances for consideration under this test include all of the defendant's circumstances '*other than those whose only reference to* [the defendant's] *conduct is that they bear on* [his] *general capacity for tolerance and self-restraint.'* This hybrid test is therefore determined to take into account relevant subjective circumstances of the defendant whilst still applying a purely objective standard of self-control. The necessity of this restriction can be

[146] *R v Goodwin* [2018] EWCA Crim 2287.
[147] *R v Acott* [1997] 2 Cr App R 94.
[148] Coroners and Justice Act 2009, s. 55(6)(c).
[149] See further Nicola Wake, 'Political rhetoric or principled reform of loss of control? Anglo-Australian perspectives on the exclusionary conduct model' (2013) 77(6) *Journal of Criminal Law* 512.
[150] Jeremy Horder and Kate Fitz-Gibbon, 'When sexual infidelity triggers murder: Examining the impact of homicide law reform on judicial attitudes in sentencing' (2015) 74(2) *Cambridge Law Journal* 307, 307 – 311.

appreciated when the defence of loss of control is considered alongside that of diminished responsibility. As Parsons explains:

> 'When the Homicide Act 1957 was enacted it was assumed that diminished responsibility was a defence which enabled those with a mental disability to be partially excused of murder on account of their disability. In contrast, the defence of provocation was for those who had full capacity but who would be partially excused because they had [not] met the standard of self-control to be expected of them.'[151]

Mindful that the previous defence of provocation was replaced by loss of control, the new defence is clearly concerned with defendants with ordinary capacities of self-control and is available only in circumstances where that ordinary capacity for self-control might have been overwhelmed in other people also. Thus, to introduce the defendant's subjective characteristics *as they relate to restraint and self-control* would be to dilute the very aim of the loss of control defence, albeit such characteristics may remain nonetheless relevant in relation to the qualifying trigger.[152] Rather, where it is claimed that some subjective characteristic such as mental illness or disease was a contributor to that loss of control, the more appropriate defence is diminished responsibility, automatism or insanity, depending upon the particular circumstances and the degree of loss of self-control.[153]

*

It is self-evident that the eponymous defence of "loss of control" speaks directly to the capacity for ordinary self-control that is presumed to exist for all adult defendants. However, whereas this capacity may also be undermined as a result of mental illness or abnormality for the purposes of the defences of insanity, automatism or diminished responsibility, loss of control is concerned with situations where the circumstances are

[151] Simon Parsons, 'The loss of control defence – Fit for purpose?' (2015) 79(2) *Journal of Criminal Law* 94, 99.
[152] Contrast *R v Rejmanski* [2017] EWCA Crim 2061 and *R v Wilcocks* [2016] EWCA Crim 2043.
[153] See further Sian Dickson and Elizabeth Stuart-Cole, 'Mentally relevant? When is a loss of control attributable to a mental condition?' (2018) 82(2) *Journal of Criminal Law* 117.

such that *anybody's* ordinary capacity for self-control might reasonably be overpowered. That being said, it is further possible that the defence may also impact upon the presumed capacity for responsiveness to reasons. As the Court of Appeal provides in *R v Jewell*,[154] loss of control may also be taken to mean a 'loss of the ability to act in accordance with considered judgment or a *loss of normal powers of reasoning*.'[155] Thus, it may be appreciated how the defence of loss of control in fact speaks to both of the volitional capacities that are presumed to exist for all adult defendants.

In a similar vein, the defence of diminished responsibility explicitly references the relevant capacities that must be impacted upon in order for the defence to be available – specifically, the defendant's ability to understand the nature of their conduct, to form a rational judgment, and to exercise self-control. Here, the latter two abilities are clearly in parallel with the presumed capacities for reasons responsiveness and ordinary self-control. Meanwhile, the first ability of understanding the nature of one's conduct parallels with the capacity for understanding the nature and consequences of one's actions which underlies the second limb of the hybrid approach to *mens rea* defended in this thesis.

Whereas the present thesis once again readily subsumes the extant defences of diminished responsibility and loss of control within the broader concept of *mens rea*, there is one aspect of the loss of control defence which would need to be approached slightly differently following a strict application of the present thesis. Specifically, the first qualifying trigger whereunder the defendant's loss of control is attributable to fear of serious violence follows an entirely subjective approach, inquiring whether or not the defendant did in fact possess such fear. Conversely, the second qualifying trigger consisting of things said or done which constitute extremely grave circumstances and induce a *justifiable* sense of being wronged, invokes an objective test taking into consideration the defendant's subjective circumstances.[156]

[154] *R v Jewell* [2014] EWCA Crim 414.
[155] *Ibid*, [23].
[156] Laura McGowan, 'Criminal Law Legislation Update' (2011) 75(1) *Journal of Criminal Law* 4, 6; Carol Withey, 'Loss of control, loss of opportunity?' (2011) 4 *Criminal Law Review* 263, 273 – 274; citing Law Commission, *Partial Defences to Murder* (Law Com No 290, 2004), 47.

It is submitted that, in keeping with the broader rejection of pure subjectivity advocated throughout this thesis, the first qualifying trigger consisting of a fear of serious violence should similarly be qualified as a *justifiable* fear of serious violence. This proposal is elucidated more fully in section 11.3.7, below, where a similar recommendation is made with regards to the defence of self-defence. Thus, it would be necessary to demonstrate objective grounds for the defendant's fear of serious violence, albeit taking into consideration any relevant subjective circumstances – such as the defendant's significantly inferior size or strength compared with their assailant. Given that the foundations of both this qualifying trigger for loss of control and the defence of self-defence is an apprehension of violence, loss of control may remain available to provide a *partial* defence where defendants have responded to a threat with their own *unreasonable* violence, which would otherwise preclude access to the total defence of self-defence. Similarly, whereas self-defence envisages responding to some *immediate* violence, loss of control may be pleaded in response to some fear of serious violence which manifests over time, for example, such as in the context of a long-term abusive relationship.[157]

### *11.3.7. Self-Defence*

Self-defence is arguably one of the most commonly well-known legal defences, albeit the defence itself extends beyond its popular understanding and may be pleaded not only when an individual acts in defence of themselves against a violent attack, but also when they act in defence of others, to protect their own property from destruction, to prevent the commission of a crime, and to assist in effecting a lawful arrest.[158] In this respect, the defence is better understood as one of "private defence", although the nomenclature of self-defence is widely used to cover all of these particular iterations. The defence is substantially governed at common law; however, the right to use reasonable force in the prevention of a crime is further contained in section 3 of the Criminal Law Act 1967, whilst a similar right in the defence of property is contained in section 5 of the Criminal Damage Act 1971. Furthermore, the substantive right to defence of the self and property

[157] See further Allison Wu, 'Going full circle: Gender and the "loss of control" defence under the Coroners and Justice Act 2009' (2019) 1(1) *Rule of Law Journal* 46, 50; citing Jeremy Horder, 'Reshaping the subjective element in the provocation defence' (2005) 25(1) *Oxford Journal of Legal Studies* 123; *Dawes* [2013].

[158] Michael Allen, *Criminal Law* (14th ed. Oxford University Press 2017), 215.

are partially codified under section 76 of that Act,[159] albeit the common law rights continue to be substantively developed in jurisprudence.[160]

Self-defence (or private defence) currently consists of two elements which together comprise an entirely subjective test with an objective test that considers subjective elements.[161] First, the defendant must have believed that the use of force was necessary, either because he, his property or another was subject to an actual or threatened attack or because he honestly believed that to be so.[162] This applies an entirely subjective test according to the facts as the defendant actually believed them to be; therefore, a defendant may still avail themselves to the defence of self-defence where they have mistakenly and even unreasonably believed an attack or threat thereof to be happening.[163] Equally, a defendant may strike pre-emptively in response to an imminent attack and is not expected to wait until he is first hit;[164] however, the apprehended attack must be imminent.[165]

Second, the degree of force that the defendant uses must be reasonable in the circumstances; this applies an objective test, albeit the test is applied having regard to the circumstances as the defendant understood them to be.[166] Section 76 of the Criminal Damage Act 1971 further clarifies that the defendant is not under any duty to retreat from aggression before utilising self-defence although his ability to have done so may be taken into consideration, whilst reasonable force must be proportionate in the circumstances. However, it is particularly notable that section 76(5A) provides that the degree of force used in circumstances where an attacker is breaking into the defendant's home will be regarded as unreasonable if it was *grossly disproportionate*, suggesting that a more disproportionate degree of force is permitted within the household invasion situation.[167] That being said, it is clear in any context that the degree of force that is deemed to be

[159] As amended by the Legal Aid, Sentencing and Punishment of Offenders Act 2012 and the Crime and Courts Act 2013.
[160] See *R v Keane* [2010] EWCA Crim 2514, [6].
[161] Nicola Monaghan, *Criminal Law Directions* (6th ed. Oxford University Press 2020), 383 – 385.
[162] See *R v Palmer* [1971] AC 814, 832; *R v Oatridge* (1992) Cr App R 367, 370.
[163] *R v Williams (Gladstone)* (1984) 78 Cr App R 276; Criminal Damage Act 1971, ss. 76(3) & (4).
[164] *Beckford* [1988].
[165] *Devlin v Armstrong* [1971] NILR 13.
[166] *R v Owino* (1996) 2 Cr App R 128, 134; *Director of Public Prosecutions v Armstrong-Braun* [1999] Crim LR 416; James Slater, 'Making sense of self-defence' (1996) 5(2) *Nottingham Law Journal* 140, 140 – 146.
[167] See *R v Ray* [2017] EWCA Crim 1391.

reasonable and proportionate is not to be evaluated entirely in the 'calm analytical atmosphere of the courtroom… with the benefit of hindsight' but in the context of the brief moments within which a defendant must respond to a perceived imminent threat.[168]

*

There is no singularly accepted rationale behind the defence of self-defence, nor the peculiarity that self-defence may be pleaded in defence to the charge of murder whereas other personal defences such as duress and necessity may not. The argument from forfeiture suggests that an aggressor forfeits the protection of some of their own rights when they attack another, entitling that other to respond in self-defence. Alternatively, a consequentialist approach justifies self-defence because the consequences of injury or death to the aggressor are "preferable" to the same consequences being suffered by an innocent victim. Further alternatively still, self-defence may be considered through the lens of competing rights of the attacker and victim, especially when the victim ends up killing their attacker in self-defence.[169] The current thesis proposes that self-defence may, again, be appreciated by reference to the three mental capacities underlying *mens rea*.

Specifically, it is submitted that the defence of self-defence operates principally by virtue of the capacity for ordinary self-control being overpowered, and secondarily by virtue of the capacity for reasons responsiveness being undermined. These connections are drawn in particular from the requirements that the defendant's own violence is both necessary in response to a (perceived) threat, and that the threat is imminent.[170] The argument follows that in the circumstances to which self-defence is restricted – *i.e.,* where there is a threat of imminent violence – the capacities of self-control and reasons responsiveness of even the hypothetical ordinary reasonable man may be so overwhelmed that responding with violence in defence is not only a reasonable response for any ordinary

[168] *Attorney-General for Northern Ireland's Reference (No. 1 of 1975)* [1977] AC 105, 138.
[169] See further Fiona Leverick, *Killing in Self-Defence* (Oxford University Press 2006), Ch. 3; Sangero (2006), Ch. 1.
[170] See further Alan Norrie, *Crime, Reason and History: A Critical Introduction* (3rd ed. Cambridge University Press 2014), 279 – 284.

individual to make, but it is even an expected, typical, and inescapably human response. The point may be illustrated through a series of hypothetical questions:

- *If a person is randomly attacked on the street, is it reasonable to expect them to stand passively and receive blows?*
- *If somebody goes to damage another's vehicle with a baseball bat, is it reasonable to expect the owner of that vehicle to stand by and watch the destruction of their property?*
- *If a person witnesses their relative or friend being attacked by another, is it reasonable to expect that witness to stand by and watch?*

In each instance the answer to the question must be negative. Crucially, however, this assertion is not made because of any moral claim that acting in defence is reasonable *per se*, (whatever the veracity of such a claim may be); rather, that acting in defence is reasonable because it is the natural, expected response of even the hypothetical ordinary reasonable man. In other words, in any of the hypothetical circumstances described above, nobody is expected to stand by whilst they, their loved ones or their property are injured by another; save for there being some realistic opportunity to escape, virtually everybody *simply would* respond by attempting to mount a defence. Were the law to disallow a defence of self-defence, it would unreasonably be expecting people to act entirely contrary to the eminently natural response to try and defend oneself, one's loved ones and property in the face of danger. This is not a realistic or practical expectation that the law could ever demand of people.

As stated above, the critical elements of the defence in this respect are the necessity of responding to a threat of violence and the imminence of that threat; the defence is so confined to circumstances where any individual might reasonably be expected to respond with "fight or flight". One of the most well-documented bodily responses to both physical and psychological emergencies such as violence and fear is a release of the hormone adrenaline. Adrenaline increases blood flow to the muscles; restricts blood flow to the skin and promotes clotting against physical trauma; releases metabolic fuels such as glucose; and stimulates respiration and heartrate. Moreover, 'from a psychological point

of view, adrenaline intensifies emotional experiences and increases… "reservoirs of power," exerting antifatigue and energizing effects.'[171] Thus, the circumstances where the defence of self-defence becomes available are such that the individual defendant will typically be faced with a fight-or-flight response, with their body and psychology modified by adrenaline to reflect that situation.

Support for this perspective may be gleaned from some of the earlier writings regarding the defence of self-defence. For example, in his seminal *Commentaries on the Laws of England*, Blackstone clearly draws a connection between a person's natural and human response to defend themselves, their property and loved ones from a threat of violence, and the reasonable user of force that is permitted when *necessary* in the face of that *imminent* threat. Indeed, Blackstone considers the right to defend oneself as the 'primary law of nature' which cannot be taken away by the 'law of society.'[172] As he writes:

> '[T]he law in this case *respects the passions of the human mind*, and (when external violence is offered to a man himself, or those to whom he bears a near connection) makes it lawful in him to do himself that immediate justice *to which he is prompted by nature, and which no prudential motives are strong enough to restrain.*'[173]

It is submitted that Blackstone's references to the "passions of the human mind" and responding in such a way that "no prudential motives are strong enough to restrain" is a reflection of the manner in which the defence of self-defence operates by reference to the capacities for ordinary self-control and reasons responsiveness. The defence is so restricted as to be available in circumstances where those capacities might fairly and reasonably be expected to be overwhelmed in the hypothetical ordinary reasonable person.

*

[171] David S. Goldstein, *Adrenaline and the Inner World: An Introduction to Scientific Integrative Medicine* (Johns Hopkins University Press 2006), 6 – 7.
[172] Blackstone (1875), 2 – 4.
[173] *Ibid*. (emphasis added).

Here, again, whilst the extant defence of self-defence may be conceptualised within *mens rea* by reference to the three underlying mental capacities, the implications of the present thesis suggest one particular amendment to the extant law. Specifically, the first limb of the current approach to self-defence is entirely subjective in asking whether the defendant actually apprehended immediate violence, even where that apprehension was both unreasonably mistaken. However, this (partially) rests the question of the defendant's criminal responsibility on their entirely subjective state of mind, whilst the present thesis has rejected such reliance on pure subjectivity. Instead, it is proposed that the defendant's apprehension of violence must be reasonable in the particular circumstances. Like the second limb of the extant defence, this would be a hybrid test which applies an objective standard – whether or not fear of violence was reasonable or justified – whilst taking into consideration the subjective circumstances of the defendant, including the situation as it presented itself to them. A similar recommendation is made regarding the corresponding qualifying trigger of fear of serious violence for the defence of loss of control, above.

As with many areas of *mens rea* explored throughout this thesis, the law has historically taken different approaches towards the subjectivity or objectivity of the first limb of the defence of self-defence, currently landing on a purely subjective approach in *R v Williams (Gladstone)*. In *R v Weston* in 1879,[174] however, the court considered that self-defence was available where the defendant used force 'against serious violence or in the *reasonable* dread of it.'[175] Similarly, in *R v Rose*,[176] in which a defendant shot and killed his violent father in the mistaken belief that he was about to attack his mother, the defendant was acquitted because 'at the time he fired that shot he honestly believed, *and had reasonable grounds for the belief*, that his mother's life was in imminent peril, and that the fatal shot which he fired was absolutely necessary for the preservation of life.'[177] Indeed, there is a wealth of authority preceding *Williams (Gladstone)* suggesting that any

[174] *R v Weston* (1879) 14 Cox CC 346.
[175] *Ibid.*, 351 (emphasis added).
[176] *R v Rose* (1884) 15 Cox CC 540.
[177] *Ibid.*, 541 (emphasis added).

mistake as to the apprehended violence must be reasonable for the purposes of self-defence.[178]

What is more, the decision to reform such a long-standing legal position from the judicial bench may be criticised from a number of directions. Firstly, notwithstanding the aforementioned authorities, the Court of Appeal in *Williams (Gladstone)* proclaimed that it had always been the case at common law that a defendant is to be judged by the honesty of his mistaken belief, whether or not it was reasonable. As Funk rightly criticises, the Court was 'either unaware of, or ignored, the English common law's earlier pronouncements that the reasonableness of defensive force was to be judged objectively.'[179] Secondly, the rationale in *Williams (Gladstone)* was substantively built upon the prior decision of *Director of Public Prosecutions v Morgan*[180] but, whereas the House of Lords in that case confirmed that mistakes raised in the support of defences *are* bound by a requirement of reasonableness, the Court of Appeal in *Williams (Gladstone)* found this not to be required for self-defence. The Court achieved this by reading the word "unlawful" into the *actus reus* of the substantive offence charged and 'by artificially making, via this term, the absence of defences part of the *actus reus*, the court was able to apply *Morgan* and to dispense with the requirement of reasonableness.'[181]

The approach in *Williams (Gladstone)* thereby fundamentally misconceives the general principle espoused by *Morgan*,[182] whilst the artificial reliance on reading "unlawfulness" into the *actus reus* of offences has been criticised for being tautological.[183] Moreover, particular arguments are raised in relation to some of the consequences of the purely subjective approach in *Williams (Gladstone)*. Specifically, as those criticisms relate to self-defence, it is argued that the purely subjective first limb of the defence gives insufficient consideration to the victim who is entirely innocent in cases of mistaken self-

[178] See *R v Foster* (1825) 1 Lewin 187; *R v Smith* (1837) 8 Car & P 158; *R v Chisam* (1963) 47 Cr App R 130; *R v Fennell* [1971] 3 All ER 215; *Palmer* [1971]; *Albert v Lavin* [1982] AC 546.
[179] T. Markus Funk, *Rethinking Self-Defence: The "Ancient Right's" Rationale Disentangled* (Bloomsbury Publishing 2021), 191.
[180] *Morgan* [1976].
[181] I. H. E. Patient, 'Mistake of law – A mistake?' (1987) 51(3) *Journal of Criminal Law* 326, 333.
[182] Simester (1992), 300.
[183] *Ibid.*, 301; Patient (1987), 334; *Albert* [1982], 561.

defence.[184] Meanwhile, including objective requirements that any such apprehension of violence be based upon reasonable grounds would require and encourage people to more carefully assess situations before resorting to violence.[185] It has been further posited that the extant law breaches the Article 2 right to life of innocent victims in cases of mistaken self-defence as, whereas the law may excuse killing another even upon an *unreasonable* mistaken belief, Article 2 requires that life may only be taken where "absolutely necessary", which is not satisfied in cases of unreasonable mistaken self-defence.[186]

It is notable that the law in Scotland already requires belief in imminent danger to be reasonable for the purposes of self-defence;[187] indeed, England occupies something of a minority position in accepting purely honest but nonetheless unreasonable mistakes as a foundation for the defence.[188] Consequently, this thesis supports a return to the position before *Williams (Gladstone)*, specifically whereby an apprehension of imminent violence must be supported by reasonable grounds under the first limb of the applicable test. It is submitted that such a reform should not have any effect in cases where a violent attack has already begun; patently there are reasonable grounds to apprehend violence in such a case. Rather, this amendment will place an additional requirement of reasonable prudence in situations where a defendant acts in self-defence to pre-empt an imminent attack. Where such an apprehension of violence lacks any reasonable grounds upon which it might have been formed, the defendant would consequently have no recourse to self-defence. This would further realign the reasonableness requirements of self-defence with the defences of duress and necessity under which the defendant's perceptions and reactions are adjudged by the reasonableness standard.

---

[184] George P. Fletcher *Rethinking Criminal Law* (Oxford University Press 2000), 689 – 690; Andrew Ashworth, 'Case comment: *Andronicou and Constantinou v Cyprus*' (1998) *Criminal Law Review* 823.
[185] Sangero (2006), 286 – 287; Simester (1992), 309.
[186] See further Fiona Leverick, 'Is English self-defence law incompatible with Article 2 of the ECHR?' (2002) (May) *Criminal Law Review* 347; John C. Smith, 'The use of force in public or private defence and Article 2' (2002) (Dec) *Criminal Law Review* 958.
[187] See further Fiona Leverick, 'Unreasonable mistake in self-defence: *Liester v HM Advocate*' (2009) 13(1) *Edinburgh Law Review* 100, 103.
[188] Claire de Than and Jesse Elvin, 'Mistaken private defence: The case for reform' in Reed A. and Bohlander M. (eds.), *General Defences in Criminal Law: Domestic and Comparative Perspectives* (Ashgate Publishing 2014), 143.

### *11.3.8. Duress and Necessity*

Duress by threats made by another person has long been recognised as a defence in English law, whereas duress by circumstances is a considerably more recent recognition. Whilst clearly there are differences between the two defences (*e.g.*, by its nature duress arises from something done by another person), both jurisprudence and academia largely treat the concepts as fundamentally related. As Tamblyn writes, duress by threats and circumstances have often been regarded as subject to the same tests in law; 'cases of one type are often discussed by the court when explaining the legal test for cases of the other type… [and] the case law for the two types is inextricably interwoven.'[189] One significantly overlooked distinction claimed by Tamblyn is that duress by threats may defend a person from acting against another individual or their property, whereas duress by circumstances provides a defence only to actions against another's property, but not their person. Furthermore, it is contentious whether or not a defence of necessity distinct from duress of circumstances may be found in English law. In any event, none of these defences are formally permitted to provide a defence to the charge of murder.[190]

One of the most authoritative statements of the components of both duress by threats and circumstances is provided by the House of Lords in *R v Hasan*:[191]

> 'To found a plea of duress the threat relied on must be to cause death or serious injury…[192] [and] must be directed against the defendant or his immediate family or someone close to him…[193] The relevant tests pertaining to duress have been largely stated objectively, with reference to the reasonableness of the defendant's perceptions and conduct…[194] The defence… is only available where the criminal conduct which it is sought to excuse has been directly caused by the threats which are relied upon…

[189] Nathan Tamblyn, *The Law of Duress and Necessity: Crime, Tort, Contract* (Routledge 2017), 166.
[190] Albeit, in relation to necessity see *Re A (conjoined twins)* [2001] Fam 147.
[191] *R v Hasan* [2005] UKHL 22, [21]; confirming *Graham* (1982).
[192] *Ibid*; citing *Director of Public Prosecutions for Northern Ireland v Lynch* [1975] AC 653, 679.
[193] *Ibid*; citing *R v Conway* [1989] QB 290; *R v Wright* [2000] Crim LR 510.
[194] *Ibid*; citing *Lynch* [1975], 670.

> [and] only if, placed as [the defendant] was, there was no evasive action he could reasonably have been expected to take.'

Thus, the defendant must first have acted in committing an offence because of a threat of death or serious harm to themselves, family or other people close to them, whereby their belief and perception of that threat is judged objectively according to the standard of reasonableness – *i.e.,* the belief in a threat of death or serious harm must have been reasonable to hold in the circumstances.[195] That threat is issued by another person in duress by threats, or arises as a result of extraneous circumstances in duress of circumstances. Mere pressure or threats of injury are not sufficient,[196] and a high threshold is maintained in relation to what will be regarded as a threat of death or serious injury.[197] Moreover, as in self-defence, that threat must be imminent such that the defendant did not have any other opportunity to escape the duress, for example, by seeking the aid of the police.[198]

With regards to the objective test of the defendant's response to those threats, it must be proven that 'a sober person of reasonable firmness, sharing the characteristics of the defendant, would [] have responded' to those threats in the same manner as the defendant; the defendant must display the 'steadfastness reasonably to be expected of the ordinary citizen in his situation.'[199] This is a hybrid test which takes into consideration the defendant's subjective characteristics on the one hand, whilst asking whether or not the sober person of reasonableness firmness imbued with those characteristics would have responded to the threats in the same way. As the hypothetical ordinary citizen encompasses a range of people of varying firmness and resistance, so mere personality traits such as frailty or timidity are not relevant characteristics to take into consideration under the reasonableness test. However, characteristics that have a clear and significant effect on an ordinary person's firmness may be taken into account, such as their age and

[195] *Graham* (1982), 241.
[196] *R v Brandford* [2016] EWCA Crim 1794.
[197] *R v Hammond* [2013] EWCA Crim 2709.
[198] *R v Hurst* (1995) 1 Cr App R 82, 93; *Hudson and Taylor* [1971].
[199] *Graham* (1982), 241.

sex,[200] pregnancy,[201] 'serious physical disability, which may inhibit self-protection, [and any] 'recognised mental illness or psychiatric condition, such as post-traumatic stress disorder leading to learned helplessness.'[202]

Notwithstanding the fact that the courts have often conflated the two,[203] there is notable contention that a defence of necessity may exist in English law distinct from duress of circumstances. Whereas duress of circumstances still requires a threat of death or serious injury, necessity could provide a defence in circumstances where such a serious threat does not exist, but where the defendant nonetheless commits some offence in order to avoid a greater evil that might otherwise have occurred. Some of the earliest English jurisprudence clearly envisaged such a defence, with such examples including raising down a house to prevent the spread of fire, permitting a felon to escape from a burning prison, or jettisoning cargo in order to save a vessel during a storm.[204] By the 19th Century, however, the law appeared to have become more equivocal; the decision of *R v Dudley and Stephens*[205] explicitly rejected that necessity could provide any defence to murder, but further appeared to doubt the veracity of the defence in general.

A number of 20th Century road traffic cases offered *obiter dicta* suggesting that the availability of a defence of necessity 'to the extent that it exists must depend on the degree of the emergency or the alternative danger to be averted.'[206] Moreover, a number of decisions appeared clearly to be reached through the application of a doctrine of necessity, even though the same was not necessarily accepted explicitly.[207] However, there appeared concurrently a number of explicit rejections of any defence of necessity in English criminal law, not least from Lord Denning MR who commented that permitting a defence of necessity would 'open a door which no man could shut… [and] would be an excuse

[200] *R v Ali* [1989] Crim LR 736.
[201] *R v GAC* [2013] EWCA Crim 1472, [33].
[202] *R v Bowen* (1996) 2 Cr App R 157, 157; see further David Cowley, 'Defence of duress – The objective test' (1997) 61(2) *Journal of Criminal Law* 178.
[203] For example, see *Conway* [1989]; *R v Martin* (1989) 88 Cr App R 343.
[204] Allen (2017), 206; citing *Moore v Hussey* (1609) Hob 96; *Mouse's Case* (1620) 12 Co Rep 63.
[205] *R v Dudley and Stephens* (1884) 14 QBD 273.
[206] *Woods v Richards* (1977) 65 Cr App R 300, 303;
[207] For example, see *Gillick v West Norfolk and Wisbech AHA* [1986] AC 112; *Re F (Mental Patient: Sterilisation)* [1990] 2 AC 1 (per Lord Bandon, but contrast with Lord Goff).

for all sorts of wrongdoing.'[208] The defence of necessity was recognised explicitly and extraordinarily in *Re A (Conjoined Twins)*, which concerned the surgical separation of conjoined twins in circumstances which would certainly kill one twin, but where both were sure to die without any such surgical intervention.[209] The Court of Appeal provided that a defence of necessity was available where the otherwise criminal act is necessary to avoid an 'inevitable and irreparable evil,' amounts to no more than is 'reasonably necessary for the purpose to be achieved,' and the evil created by the criminal act must not be disproportionate to the evil that is being avoided.[210]

Whereas the Court of Appeal in *Re A (Conjoined Twins)* strained to confine the defence strictly to the unique facts of that particular case, the application of the doctrine may be found in a number of instances, with quintessential examples being where traffic laws are breached in order to rush somebody to hospital,[211] or where medical treatment is justified in order to avoid certain physical and mental suffering.[212] Crucially, in this respect, a plea of necessity must appeal to *extraneous circumstances* as the cause of the commission of the offence; for example, the cultivation of marijuana in order to treat the defendant's pain does not qualify, because the cause of the commission of the crime – *i.e.,* pain – was not an extraneous circumstance.[213] The unifying feature in accepted examples of necessity therefore appears to be that the evil being avoided is that of significant pain and suffering caused by extraneous circumstances, albeit not reaching the degree of threat to life or serious injury required under duress of circumstances.

*

[208] *London Borough of Southwark v Williams* [1971] 2 All ER 175, 179; *Buckoke v GLC* [1971] Ch 655, 668; *R v Kitson* (1955) 39 Cr App R 66.
[209] See further Jonathan Rogers, 'Necessity, private defence and the killing of Mary' (2001) (Jul) *Criminal Law Review* 515; see also *R v Bournewood Community and Mental Health NHS Trust* [1998] 3 All ER 289, 300.
[210] *Re A (conjoined twins)* [2001], 240.
[211] *Director of Public Prosecutions v Pipe* [2012] EWHC 1821.
[212] For example, *Gillick* [1986]; *R v West Berkshire Health Authority* [1989] 2 AC 1; *R v Bourne* [1939] 1 KB 687.
[213] *R v Quayle* [2005] EWCA Crim 1415; David Ormerod, 'Necessity of circumstance' (2006) (Feb) *Criminal Law Review* 148.

In a similar manner to self-defence, it may readily be appreciated how the defences of duress and necessity operate by reference to the mental capacities underlying *mens rea*, and the presumed capacities for reasons responsiveness and ordinary self-control in particular. Each of the defences arise *only* where there is a threat of death or serious injury (in the case of duress) or of significant pain and suffering from an extraneous source (in the case of necessity) directed towards either the defendant, his loved ones, or those directly under his care. By virtue of the operative legal tests, the defences apply in circumstances where the steadfastness or firmness of ordinary reasonable people would similarly and fairly be regarded as being overwhelmed. In such circumstances, and faced with grave dangers to themselves or loved ones, even reasonable people do not operate with the ordinary capacities to be responsive to reason and exercise self-control. Presented again in the form of a number of questions:

- *If a person is threatened with a knife unless they steal something from a shop, is it reasonable to expect them to stand steadfast and refuse because theft is illegal?*
- *If a fire is spreading through a building, is it reasonable to expect rescuers not to break windows to help people inside?*
- *If a person's child, brother or close friend suffers a significant and life-threatening injury, is it reasonable to expect that person to drive perfectly at the speed limit on a clear road en route to the hospital?*

Once again, in the particular circumstances where duress and necessity are potentially applicable, it may readily be appreciated that even the hypothetical reasonable person does not operate and make decisions with their ordinary capacities of responding to good reason and exercising self-control. To the individual breaking speed laws to rush a loved relation to hospital, or breaking windows to rescue them from a burning building, ordinary considerations such as the illegality of their immediate actions do not arise. As Spain writes, the 'Aristotelian view of duress, as a factor affecting the voluntariness of an act and overbearing the will of the defendant, dominates the common law system today.'[214]

[214] Eimear Spain, *The Role of Emotions in Criminal Law Defences: Duress, Necessity and Lesser Evils* (Cambridge University Press 2011), 149.

This link is drawn through the language that has rather consistently attached to the defences of duress and necessity. For example, the Court of Appeal in *Graham* describes how 'in duress the words or actions of one person [or situation] break the will of another.'[215] Similarly, the House of Lords in *R v Howe*[216] described the defence of duress as a 'concession to human frailty' under which even a considered and conscious decision to commit some criminal act is made in order to save oneself or loved ones.[217] As with self-defence, the argument being made is not the moral claim that acting under duress or out of necessity is fundamentally reasonable *per se*, (whatever the veracity of such a claim); rather, that acting under duress or out of necessity is reasonable because it is the natural, expected, even inevitable response of the hypothetical ordinary reasonable person. In this regard, the law does not punish a person for acting in precisely the way that would fairly and reasonably be expected of any other reasonable member of society, in the circumstances to which duress, necessity and, indeed, self-defence are restricted.

Discussed at greater length in section 11.2 of this thesis, above, duress, necessity and self-defence may together be referred to as defences arising out of "circumstances of compulsion." In this regard, the defences are available when the circumstances giving rise thereto are such that *any reasonable person* is reasonably expected to succumb to those circumstances. Therefore, the claim is not made that it is the subjective *characteristics of the particular defendant* such as some mental illness or other deficiency which diminishes their relevant capacities of responding to reason or exercising ordinary self-control. Rather, these defences point to the *characteristics of the particular threat* within the aforementioned circumstances of compulsion which, it is argued, are of such gravity that they overwhelm the volitional capacities of any reasonable person. Put differently, the defendant's "choice" to commit what would otherwise be a criminal offence under these defences is no genuine choice at all – their actions are "morally involuntary." This interpretation of defences of compulsion has a strong pedigree in jurisprudence extending back to the seminal work of William Blackstone.

[215] *Graham* (1982), 241.
[216] *R v Howe* [1987] AC 417.
[217] *Ibid.*, 435.

## *11.4. A New Defence of Addiction*

The present chapter of the thesis identifies formal legal defences as claims affecting one or more of the three mental capacities regarded as fundamental to ascriptions of responsibility. That is to say, each recognised legal defence may be related to the way in which certain characteristics or circumstances are understood to sufficiently diminish, undermine or abrogate entirely one or more of the capacities for recognising and responding to reason, ordinary self-control, and appreciating the nature and consequences of one's actions and decisions. From this point, the thesis can fulfil the practical role of both rationalising why certain defences are accepted as such and, crucially, identifying new potential defences. In essence, new defences might be recognised wherever characteristics or circumstances similarly diminish, undermine or abrogate entirely one or more of the three aforementioned capacities.

### *11.4.1. Crime and Addiction*

Addiction is one such characteristic or circumstance which may prove eminently appropriate for consideration as a legal defence, considering each of the prevalent correlations between substance abuse / dependency and crime,[218] and the potential for addressing addiction as a cause of criminal behaviour within a rehabilitative setting. To consider some statistics, one Europe-wide study places the lifetime prevalence of illicit drug use at 60.6% of European prison populations, contrasted against 29% of the European population generally; prevalence of illicit drug use was similarly high during the last year (57.4%), last six months (43.3%) and last month (60.5%) prior to incarceration.[219] In a further study of UK prison populations, 57.5% were identified as having drug (non-alcohol) dependency, 29% were identified as alcohol dependent, and 20% were poly-substance dependent (drugs and alcohol).[220] The Alcohol Health Alliance

[218] For an in-depth review, see Trevor Bennett and Katy Holloway, *Understanding Drugs, Alcohol and Crime* (Open University Press 2005).

[219] Frank C. van de Baan, Linda Montanari, Luis Royuela and Paul H. H. M. Lemmens, 'Prevalence of illicit drug use before imprisonment in Europe: Results from a comprehensive literature review' (2021) 29(1) *Drugs: Education, Prevention and Policy* 1, 6 – 8.

[220] Jane L. Ireland and Pauline Higgins, 'Behavioural stimulation and sensation-seeking among prisoners: Applications to substance dependency' (2013) 36(3-4) *International Journal of Law and Psychiatry* 229, 231; see also Nicola Singleton, Michael Farrell and Howard Meltzer, 'Substance misuse among prisoners in England and Wales' (2003) 15(1-2) *International Review of Psychiatry* 150, 150.

UK estimates that up to one million criminal offences each year are associated with alcohol consumption alone, including just over one-half of all violent crimes.[221]

At least two broad factors may be appreciated for their role in contributing to addiction generally – neuropsychological factors and psychosocial factors. Beginning with the former, brain disease models of addiction focus on the role of the brain and mechanisms therein which result in the symptoms related to addiction behaviours.[222] For example, the "Incentive-Sensitisation Theory" of addiction posits that repeated exposure to an addictive substance or activity produces hypersensitivity thereto in the mesocorticolimbic regions of the brain, heightening the incentive salience of the substance or activity which, through repeated exposure, results in incremental neuroadaptations 'rendering it increasingly and perhaps permanently hypersensitive' to the addictive stimuli.[223] The "Opponent-Process Theory" considers addiction through the lens of dysregulation in homeostasis, whereby the initial pleasure of an addictive substance or activity is followed by negative counter-experiences such as withdrawal as the brain attempts to return to homeostatic balance.[224] Neurobiological adaptation results in a dysregulation of homeostasis over time as tolerance results in smaller positive effects and withdrawal results in larger negative affect, creating a cycle of binge intoxication, negative affect, and preoccupation with withdrawal which resumes the cycle.[225]

The "Cue-elicited Craving Model" follows research demonstrating how addiction-related cues induce the same neurochemical and behavioural activities as the addictive substance or activity itself.[226] The addictive stimulus obtains stronger salience within the anterior

[221] Alcohol Health Alliance UK, 'Measuring up: The state of the nation' (Alcohol Health Alliance UK 2017), 5 & 8 <https://ahauk.org/wp-content/uploads/2017/12/7119-AHA-10-year-anniversary-report.pdf> accessed 5 October 2022.
[222] Francesca Mapua Filbey, *The Neuroscience of Addiction* (Cambridge University Press 2019), 9.
[223] Terry E. Robinson and Kent C. Berridge, 'The neural basis of drug craving: An incentive-sensitization theory of addiction' (1993) 18(3) *Brain Research Reviews* 247, 247; Terry E. Robinson and Kent C. Berridge, 'The incentive sensitization theory of addiction: Some current issues' (2008) 363(1507) *Philosophical Transactions of the Royal Society: Biological Sciences* 3137.
[224] Richard L. Solomon and John D. Corbit, 'An opponent-process theory of motivation' (1974) 81(2) *Psychological Review* 119.
[225] George F. Koob and Michel Le Moal, 'Drug abuse: Hedonic homeostatic dysregulation' (1997) 278(5335) *Science* 52.
[226] Peter W. Kalivas and Nora D. Volkow, 'The neural basis of addiction: A pathology of motivation and choice' (2005) 162(8) *American Journal of Psychiatry* 1403.

cingulate and amygdala (associated with motivation and emotion respectively), whilst interoceptive and memory processes catalyse activity in the insula and hippocampus, triggering dopamine release which encodes the conditioned associations between the substance / activity and environmental cues. Where most such theories converge is in the view that addiction "hijacks" the brain's dopaminergic reward system which produces the sensations of satisfaction, enjoyment or pleasure associated with activities in which a person engages, underlying the motivational processes which cause people to crave and, ultimately, engage in those activities again in the future.[227] When this motivational reward system becomes pathologically engaged by some addictive substance or activity and its associated cues, addictive behaviours such as craving, withdrawal and preoccupation emerge, and people find themselves increasingly unable to volitionally resist the substance or activity which engages the dopaminergic system.

Beyond the neuropsychological processes in the brain which offer a partial explanation for addiction and associated behaviours, a large number of psychosocial factors are hypothesised to contribute a significant role in addiction.[228] Briefly, children who are victims of abuse (whether physical, sexual or emotional) and who exhibit more externalising behaviours such as attention deficit, hyperactivity and oppositional defiance are at higher risk of addiction later in life. Similarly, a number of personality types and childhood temperaments are associated with later addiction, including antisocial behaviour, deviance, impulsivity, mood instability and social withdrawal. The influence of both family and peers may also be a significant contributing factor throughout life from childhood to adulthood. Parental approval of, or engagement in, substance use and / or

[227] See further Francesca M. Filbey and Samuel K. DeWitt, 'Cannabis cue-elicited craving and the reward neurocircuitry' (2012) 38(1) *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 30; Francesca M. Filbey, Eric D. Claus and Kent E. Hutchinson, 'A neuroimaging approach to the study of craving' in Adinoff B. and Stein E. A. (eds.), *Neuroimaging in Addiction* (John Wiley & Sons 2011); Francesca M. Filbey, Joseph P. Schacht, Ursula S. Myers, Robert S. Chavez and Kent E. Hutchinson, 'Marijuana craving in the brain' (2009) 106(31) *Proceedings of the National Academy of Sciences* 13016; Nora D. Volkow, Joanna S. Fowler, Gene-Jack Wang, Ruben Baler and Frank Telang, 'Imaging dopamine's role in drug abuse and addiction' (2009) 56(1) *Neuropharmacology* 3; Nora D. Volkow, Joanna S. Fowler and Gene-Jack Wang, 'Role of dopamine in drug reinforcement and addiction in humans: Results from imaging studies' (2002) 13(5) *Behavioural Pharmacology* 355.

[228] See Monica C. Skewes and Vivian M. Gonzalez, 'The biopsychosocial model of addiction' in Miller P. M., Kavanagh D. J., Kampman K. M., Bates M. E., Larimer M. E., Petry N. M., DeWitte P. and Ball S. A. (eds.), *Principles of Addiction: Comprehensive Addictive Behaviors and Disorders: Volume 1* (Academic Press 2013), 64 – 68.

abuse is associated with later addiction; whilst maintaining peers who are more deviant / less conventional similarly enhances the risk of addiction throughout life. Equally, the absence of family / peers (*i.e.,* social exclusion) can similarly contribute to addiction. Life experiences which result in classical and operant conditioning may result in addiction behaviours, such as relying on alcohol or other substances as a form of stress relief to the point that using substances becomes a conditioned response to stressors. In this regard, of course, stress itself is another significant factor which contributes to addiction; equally, circumstances resulting in high stress / social exclusion, such as unemployment and poverty, are associated with addiction. Outcome expectancies regarding engagement with addictive substances / activities, and self-efficacy in the ability to so engage, are further contributing factors.[229]

Exploring some of these factors in closer detail, stress is perhaps one of the most obvious contributing factors potentially leading to addiction behaviours.[230] Stress ordinarily impacts *inter alia* on the hypothalamus-pituitary-adrenal ('HPA') axis, one of the body's key stress response mechanisms resulting in release of the stress hormone cortisol; however, exposure to chronic stress can pathologically damage this pathway resulting in psychological, metabolic and immune changes.[231] As Goeders explains, abstinence from addictive substances or activities, exposure to stressors, and addiction-related cues can repeatedly stimulate the HPA axis, producing craving for, and relapse to, that addictive agent. Consequently, 'these cues trigger the HPA axis unpredictably and without warning so that the addict feels a loss of control, and the relapse to drug use helps the individual regain control over his or her HPA axis activation.'[232] Stress also interacts with the dopaminergic reward system, highlighted above. Interestingly, mild and moderate stress can actually increase dopamine transmission, such as enjoyed during extreme sports;

---

[229] *Ibid.*

[230] For a thorough overview, see Mustafa al'Absi (ed.), *Stress and Addiction: Biological and Psychological Mechanisms* (Academic Press 2007).

[231] For example, see Thomas Frodl and Veronica O'Keane, 'How does the brain deal with cumulative stress? A review with focus on developmental stress, HPA axis function and hippocampal structure in humans' (2013) 52(1) *Neurobiology of Disease* 24; Anna Gądek-Michalska, Jadwiga Spyrka, Paulina Rachwalska, Joanna Tadeusz and Jan Bugajsk, 'Influence of chronic stress on brain corticosteroid receptors and HPA axis activity' (2013) 65(5) *Pharmacological Reports* 1163; Bruce S. McEwan, 'Protective and damaging effects of stress mediators' (1998) 338(3) *New England Journal of Medicine* 171.

[232] Nicholas E. Goeders, 'The hypothalamic-pituitary-adrenal axis and addiction' in al'Absi M. (ed.), *Stress and Addiction: Biological and Psychological Mechanisms* (Academic Press 2007), 21.

however, extreme and chronic stress can have a deleterious effect,[233] such as might be experienced with unemployment, poverty, familial breakdown, abuse, homelessness and social exclusion. This research identifies stress reduction, improved coping mechanisms, and pharmacotherapies targeting the HPA axis and dopaminergic reward pathways as potential avenues for treating addiction, reducing cravings and promoting abstinence.

Social marginalisation and exclusion is identified as a significant potential contributor to addiction in its own right. Heilig, Epstein, Nader and Shaham argue, on the one hand, that social interactions can be highly stressful when antagonistic or exclusory whereas, on the other hand, positive and healthy social relationships can be one of the most important reinforcers that compete with the rewards of addiction and 'protect against the negative consequences of social stressors.'[234] In a similar vein, Scherbaum and Specka found that the long-term abuse of opiates could be predicted from a number of factors related to social relationships, including association with other addicted peers, breakdown of familial relationships, unemployment and the lack of healthy social support.[235] Again, this research further identifies the (re)establishment of healthy social relationships as a potentially powerful route to preventing and treating addiction.

### *11.4.2. Addiction Defence and Sentencing*

The criminal law in the UK currently does not recognise any general defence attached to or concerning addiction, nor does it readily consider addiction or the effects thereof within existing defences. For example, discussed above in section 11.3.4 of this thesis, the defence of insanity requires that a defendant did not know the nature and quality of their actions, whereas addiction is more readily considered as a defect of *volition* such that, even if an addict feels that they cannot help but act to satiate their addiction, they typically remain capable of understanding the nature of their actions (indeed, addicts will act in a "rational", goal-oriented way, save for the fact that it is the satiation of their addiction

[233] Michela Marinelli, 'Dopaminergic reward pathways and effects of stress' in al'Absi M. (ed.), *Stress and Addiction: Biological and Psychological Mechanisms* (Academic Press 2007), 41.
[234] Markus Heilig, David H. Epstein, Michael A. Nader and Yavin Shaham, 'Time to connect: Bringing social context into addiction neuroscience' (2016) 17(9) *Nature Reviews Neuroscience* 592, 592.
[235] Norbert Scherbaum and Michael Specka, 'Factors influencing the course of opiate addiction' (2008) 17(S1) *International Journal of Methods in Psychiatric Research* S39.

that is the goal of their behaviour). Similarly, discussed above in section 11.3.5, the defence of automatism requires a total absence of self-control such as may occur during an epileptic seizure or hypoglycaemic attack whereas, again, whilst addiction may be considered as a defect of volition, it scarcely amounts to a total abrogation of bodily control. Meanwhile, discussed in section 11.3.6, although the defence of loss of control may permit that loss of control to be gradual and / or partial, it must originate from an exogenous qualifying trigger such as fear of serious violence or some other, extremely grave words or acts which create a justifiable sense of being serious wronged, which plainly does not apply to addiction.[236]

Some limited accommodation has been made for addiction within the defence of diminished responsibility, discussed in section 11.3.6 of this thesis, above. In brief, this statutory defence provides a *partial* defence to the offence of murder *only*, reducing the charge to one of manslaughter. The defence is established when the defendant kills another whilst suffering from an 'abnormality of mental functioning' arising from a 'recognised medical condition', which 'provides an explanation' for their act of killing another, and 'substantially impaired' their ability to understand the nature of their conduct, form a rational judgment or exercise self-control.[237] As a partial defence, the "substantial impairment" of one or more of the defendant's relevant capacities may be less than that which is otherwise required for the complete defences of insanity or automatism.

Concerning addiction as a "recognised medical condition", the Diagnostic and Statistical Manual of Mental Disorders ('DSM-5-TR')[238] published by the American Psychiatric Association ('APA') recognises a number of addiction and compulsion disorders, including substance use disorder ('SUD') and addictive disorder (in relation to gambling); whilst the International Classification of Diseases ('ICD-11')[239] published by the World

[236] See further Alan Bogg and Jonathan Herring, 'Addiction and responsibility' in Herring J., Regan C., Weinberg D. and Withington P. (eds.), *Intoxication and Society: Problematic Pleasures of Drugs and Alcohol* (Palgrave Macmillan 2013).
[237] Homicide Act 1957, s. 2.
[238] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders* (5th ed. American Psychiatric Association 2013).
[239] World Health Organization, *International Classification of Diseases* (11th ed. World Health Organisation 2019).

Health Organization ('WHO') recognises various substance addictions, gambling addiction, and compulsive sexual behaviour disorders.

The case of *R v Byrne,*[240] discussed above in section 11.3.6 of this thesis, provides an early demonstration of the diminished responsibility defence in relation to a sexual psychopath who suffered from 'violent perverted sexual desires which he [found] difficult or impossible to control', albeit he could otherwise function relatively normally and was capable of rational thought, thus falling short of the technical "insanity" definition. Following medical evidence, the Court of Appeal determined both that his mental functioning was indeed abnormal, and that this arose from a recognised medical condition. The court contrasted an "abnormality of mind" with the stricter test of "defect of reason" required for an insanity defence, opining that the abnormality of mind concept is 'wide enough to cover the mind's activities in all its aspects, not only the perception of physical acts and matters, and the ability to perform a rational judgment as to whether an act is right or wrong, *but also the power to control physical acts in accordance with that rational judgment*.'[241] Further, the Court considered that the judgment of whether an abnormality of mental functioning provides an explanation for, or contributes significant to, the defendant's criminal conduct 'points to a consideration of the extent to which the accused's mind is answerable for his physical acts which must include consideration of the extent of his ability to exercise will power to control his physical acts.'[242]

The law has struggled more in incorporating the effects of alcoholism and alcohol consumption within the defence of diminished responsibility, albeit greater clarity has emerged in recent years. To begin, it is well established that the *voluntary* consumption of alcohol *alone* cannot give rise to the abnormality of mind arising from a recognised medical condition for the purposes of the diminished responsibility defence.[243] Rather, whereas alcoholism or substance addiction could establish the relevant abnormality of mind, the effects of consuming alcohol or drugs in the event would only be relevant if that consumption had caused damage to the brain or had produced an irresistible craving

[240] *Byrne* [1960].
[241] *Ibid.,* 403.
[242] *Ibid.*
[243] *R v Fenton* (1975) 61 Cr App R 261; *R v Gittins* [1984] QB 698.

such that the consumption was involuntary.[244] The House of Lords later provided a model direction in *R v Dietschmann*:[245]

> 'Drink cannot be taken into account as something which contributed to [the defendant's] mental abnormality and to any impairment of mental responsibilty arising from that abnormality. But you may take the view that both the defendant's mental abnormality and drink played a part in impairing his mental responsibility for the killing and that he might not have killed if he had not taken drink. If you take that view, then the question for you to decide is this: has the defendant satisfied you that, despite the drink, his mental abnormality substantially impaired his mental responsibility for his fatal acts, or has he failed to satisfy you of that? If he has satisfied you of that, you will find him not guilty of murder but you may find him guilty of manslaughter. If he has not satisfied you of that, the defence of diminished responsibility is not available to him.'[246]

The case of *R v Wood*[247] concerned an alcoholic who, significantly inebriated from drinking large quantities of alcohol, awoke to find another making unwanted sexual advances towards him. The defendant subsequently lost control and killed his victim by striking him repeatedly with a meat cleaver, later pleading diminished responsibility at trial on account of his alcoholism. The Court of Appeal considered the impact of *Dietschmann*, again finding that the consumption of alcohol or drugs alone could not support a defence of diminished responsibility; however, the Court did consider that the voluntary consumption of alcohol would not necessarily defeat the defence. Rather, the crucial question remained with the effect of the abnormality of mind arising from a recognised medical condition, *i.e.,* the effect of the defendant's alcoholism:

> 'If the syndrome does not constitute such an abnormality of mind, diminished responsibility based on the consumption of alcohol will fail. If,

[244] *R v Tandy* (1988) 87 Cr App R 45.
[245] *R v Dietschmann* [2003] 1 AC 1209.
[246] *Ibid.,* 1227.
[247] *R v Wood* [2008] EWCA Crim 1305.

> on the other hand, it does, the jury must then be directed to address the question of whether the defendant's mental responsibility for his actions at the time of the killing was substantially impaired as a result of the syndrome. In deciding that question the jury should focus exclusively on the effect of alcohol consumed by the defendant as a direct result of his illness or disease and ignore the effect of any alcohol consumed voluntarily. Assuming that the jury has decided that the syndrome constitutes an abnormality of mind induced by disease or illness, its possible impact and significance in the individual case must be addressed. The resolution of this issue embraces questions such as whether the defendant's craving for alcohol was or was not irresistible, and whether his consumption of alcohol in the period leading up to the killing was voluntary (and if so, to what extent) or was not voluntary, and leads to the ultimate decision, which is whether the defendant's mental responsibility for his actions when killing the deceased was substantially impaired as a result of the alcohol consumed under the baneful influence of the syndrome.'[248]

Thus, the present state of the law does recognise various addictions as relevant medical conditions contributing to a sufficient abnormality of mind and, further, permits consideration of the behavioural effects of those addictions with regards to whether they provide an explanation for, or significantly contribute to, a defendant's subsequent criminal conduct. However, the present law is most obviously limited in restricting the partial defence of diminished responsibility to the offence of murder only whereas, as highlighted at the outset of the present section of this thesis, addiction may be a significant contributing factor to a whole range of offences.

*

This thesis supports the introduction of a wider (albeit still partial) defence of addiction, loosely modelled on the approach of diminished responsibility. First, as discussed in the outset of this section of the thesis, strong correlations can readily be drawn between

[248] *Ibid.,* [41]; see also *R v Stewart* [2009] EWCA Crim 593.

various different addictions and resultant criminal behaviour, whether the object of an addiction is itself criminal (such as illegal drugs), or whether the object of an addiction is liable to lead to other criminal conduct (such as with the consumption of alcohol). Second, addiction is fuelled by a number of identifiable causes many of which lay outside of an individual's sphere of personal control. For example, childhood abuse and poverty, parental attitudes towards addictive substances and activities, and certain developed personality traits and temperaments all represent factors over which individuals have little active control during their upbringing, and yet may significantly contribute to addiction behaviours later in life. Third, addiction is something which is treatable in principle, for example, with pharmacological interventions targeting the HPA axis in response to stressors, psychotherapeutic interventions such as cognitive behavioural therapy providing people with stronger mechanisms for coping with stressors, and psychosocial approaches such as providing greater community social support structures. Thus, a consequentialist theory of responsibilty such as presented in this thesis should adopt an holistic attitude towards the issue of addiction, recognising the causes which lay outside of individual control, the relationships between addiction and criminal behaviour, and the interventions that can address addiction and thereby reduce associated crime.

Whilst it is not intended to place any hard restrictions on the types of addiction which may be considered under the proposed defence, at least three areas are identified for having particular potential relevance to subsequent criminal conduct. First and foremost, substance addictions have a clear and obvious connection with criminal behaviour; addictions to illicit drugs are liable to result in drug possession offences, whilst addictions to both illicit drugs and / or alcohol are liable to result in criminal conduct flowing from the effects these substances have on behaviour, aggression, and rational decision-making whilst under the influence. Additionally, substance addictions may be liable to result in property offences in order to fund the addiction. Second, gambling addictions are similarly liable to result in property offences in order to fund the addiction. Third, sex addictions may be liable to result in the commission of sexual offences. The existing defence of diminished responsibility requires that the relevant abnormality of mental functioning substantially impairs the defendant's capacity to understand the nature of their conduct, to form a rational judgment (*i.e.,* recognise and apply reason in their

thinking), or to exercise self-control. Equally, the proposed addiction defence would only be available to defendant's who have reached the level of clinical addiction such that one or more of these capacities is substantially impaired; the defence would therefore not be available for less severe dependencies or compulsions which do not substantially interfere with the crucial capacities for responsibility.

As discussed above, the level of impairment required for the defence of diminished responsibility is less than that required under such defences as insanity and automatism. Consequently, where these latter pleas offer a complete defence to criminal conduct, the former plea of diminished responsibility offers only a partial defence, reducing a charge of murder to one of manslaughter. On the one hand, it is proposed that the defence of addiction would apply beyond murder to cover all criminal offences as, patently, addiction behaviours may be eminently relevant to drug possession, violent, property, and sexual offences. On the other hand, it is proposed that the addiction defence would only be partial, as with diminished responsibility. In this regard, addiction rarely abrogates a person's capacities entirely; for example, addicts may make decisions that are ultimately "rational" vis-à-vis satisfying their addiction, and can similarly exhibit ordinary self-control vis-à-vis the actions required to satisfy an addiction. In the rarer circumstances where addiction behaviours resulted in a complete abrogation of capacity – (for example, where alcohol or drug consumption results in psychosis and an accompanying inability to appreciate the nature and consequences of one's actions or a total lack of ordinary self-control) – it is submitted that the more appropriate defence would be that of insanity or automatism.

The defence of diminished responsibility requires that the defendant's abnormality of mental functioning provides an explanation for, or significantly contributed to, their subsequent criminal behaviour. The same condition is proposed to be included in the defence of addiction; as such, addiction will offer no defence where the offence charged bears little or no relation to the addiction suffered. Rather, the defendant's particular addiction must provide an adequate explanation for, or substantially contribute to, the criminal charge against which they are attempting to defend. By way of example, whereas an addiction to illicit substances might potentially explain or contribute towards property

offences committed in order to fund that addiction, a sexual addiction is highly unlikely to be similarly relatable to *property* offences. As a partial defence, a successful addiction plea would reduce the offence charged to its lesser equivalent, for example, reducing murder to manslaughter, burglary / robbery to theft, wounding with intent to wounding *simpliciter*, and rape to sexual assault, *etc.* This is intended to reflect the manner in which addiction typically diminishes an individual's volition without abrogating their capacities entirely.

Finally, it is proposed that a successful defence of addiction would carry at least two sentencing implications. Primarily, any sentence ought to include a rehabilitation / treatment component to ensure that the defendant receives appropriate medical, pharmacological and / or psychotherapeutic treatment for their addiction. Further, wherever possible, it is proposed that the remaining component of sentences would prefer community services orders over incarceration for at least two reasons. First, considering the potential contribution of significant stressors to addiction behaviours, it would appear counterproductive to place a recovering addict within a highly stressful prison environment where relapse may be more likely, (not to mention the prevalent availability of drugs within prisons). Second, considering the potential contribution of social relationships to successful recovery, it is suggested that healthier social relationships might be built in a community setting as contrasted against the prison environment where the addict is not only excluded from mainstream society, but is additionally placed in an environment surrounded by other deviant offenders. To the extent that the criminal justice system aims to prevent crime in the first instance and reduce recidivism in the second instance, placing addicts within the prison population may ultimately increase recidivism and, therefore, may be inimical to these broader aims.

# 12. Verdict and Punishment

> 'Punishment is not for revenge, but to lessen crime and reform the criminal.'
>
> - Elizabeth G. Fry, 1847.[1]

The subject of punishment alone can fill volumes and, as this thesis primarily concerns the issue of responsibility, it is not necessary to conduct a complete examination of this topic. That notwithstanding, the present thesis does create some strong implications regarding the subject which are pertinent to explore. To preface, it is useful to highlight some of the key elements of the thesis that are most relevant to the topic of punishment. The chapter then proceeds to discuss punishment, focusing on the implications that the present thesis poses for the various theories of punishment which underlie the practical (and political) processes of determining and implementing different sentences for criminal offences. In particular, the thesis rejects retributivism as being inherently incompatible with the deterministic, consequentialist and capacity-based theory of responsibility that has been presented. Furthermore, the thesis advocates strongly in support of rehabilitation and deterrence as the preferable guiding principles for punishment and sentencing, whilst incapacitation remains an overarching principle of punishment that is available at all times when necessary. From this discussion emerges a general hierarchy between the remaining theories of punishment, which the thesis maps loosely onto a discussion of verdicts. In this latter respect, the thesis proposes replacing the verdict of "not guilty by reason of insanity" with a broader verdict of "not responsible", following which the court has access to some sentencing options, but not the full range of options which follow a "guilty" verdict.

[1] Elizabeth Gurney Fry, *Memoir of the Life of Elizabeth Fry: With Extracts from Her Journal and Letters: Vol. 1* (Fry K. and Creswell R. E. (eds.), Charles Gilpin and John Hatchard & Son 1847), 309.

The theory of responsibility here presented began with an exploration of how the brain makes decisions, including decisions to commit criminal acts. Multi-alternative decision field theory describes the manner in which networks of neurons representing different available decision outcomes compete towards a threshold at which point a decision is made. This description was then disambiguated to five "components" of any decision – the *what, how, when, whether* and *why* of a decision. With regards to each component, evidence was presented significantly revealing the apparent automaticity of the distinct brain networks responsible for their respective decision components; that is, the brain appears to be able to processes each of these components without the *necessary* intervention of conscious awareness, effort or control. This is not to say that consciousness *necessarily* has nothing to offer to decision-making and, indeed, its precise role and contribution remains one of the great questions of neuroscience, psychology and philosophy. Furthermore, it has been hypothesised at various points throughout this thesis that consciousness may improve decision-making by offering more time and greater mental resources (*i.e.,* energy, oxygen, *etc.*) to the different neural process involved. Thus, by consciously deliberating and concentrating on a particular problem or decision, the operation of the various decision-making networks may be improved.

The aforementioned notwithstanding, the evidence strongly indicates that each of the brain networks engaged in the respective components of any decision *can and do* operate automatically and unconsciously. Indeed, whilst the evidence can fairly be interpreted to suggest that a significant majority, if not all, decisions are ultimately the result of such automatic processes, at a very minimum it may still be said that an unknowable proportion of our decisions are so automatic. As neuronal networks representing different decision outcomes compete for supremacy, a broad range of both internal stimuli – homeostatic states, emotions, memories, *etc.,* – and external stimuli – which may be anything from the mere environment that a person finds themselves in to the purposive actions of another agent – contribute to those networks as they each recruit "evidence" in support of different decision outcomes. Thus, not only does the decision-making architecture operate automatically, but it *can and does* process goals and intentions that can fairly be said to originate entirely exogenously of the individual. By extrapolation, it follows that an unknowable proportion of criminal actions may arise as a result of the automatic

processing of initial goals or intentions, themselves which may be triggered by entirely external stimuli, resulting in consequent behaviour without the opportunity for conscious awareness or, therefore, intervention and control of that decision and behaviour.

The theory of responsibility which followed in the second part of this thesis is grounded upon three capacities and is consequentialist in nature. The capacities for being responsive to reason in the process of making decisions and for exercising self-control (but not necessarily *conscious* control) over resultant behaviour are each part of the human condition and are ordinarily exhibited by most people. That is to say, the competing networks of neurons representing different decision outcomes are indeed responsive to reasons (*i.e.,* different evidence) that are recruited in their support or opposition, whilst we all typically experience and exhibit the ability to control our physical behaviour such that it accords with our goals and intentions.[2] As such, it is perhaps appropriate that the law presumes each of us to possess these capacities within the concept of volition, unless evidence is presented to the contrary in the form of a relevant legal defence.

The capacity to appreciate the nature and consequences of one's actions is altogether more contextual; it is much less an "ordinary" capacity such as self-control which assumes a consistent and continuous role in the regulation of daily life. Rather, the extent to which an individual might appreciate the nature of their actions can vary depending upon a wide range of factors, from such intrinsic matters as a person's experience and familiarity with a given situation or problem, to such relatively trivial matters as how much that person was paying attention in the moment of making a particular decision. As such, the law does not presume that this capacity was competent at the time of a defendant's alleged offending but, as conceptualised under the present thesis, the functioning of this capacity must be proven by the prosecution within the revised concept of *mens rea*.

It is the committing of a criminal offence (*actus reus)* in the presence of these three capacities – encapsulated as they are by the concept of *mens rea* and volition – which

[2] Although, following the discussion in chapters four and five of this thesis, it may well be that our behaviour generally accords with our intentions because the experience of intentionality is an inherent component of the preparation of motor actions *and* the assessment of their success in meeting their underlying purpose or goal.

culminates in a person's criminal responsibility. That is to say, an individual may be held responsible for their actions if: they possessed the capacity to be responsive to reasons such that their actions could in principle be guided by the fact that certain conduct is prohibited by law; they possessed the capacity for ordinary self-control such that, having determined not to do some criminal act, they could consequently control their actions in concurrence with that decision; and they possessed the capacity to appreciate the nature and effects of their actions such that they would be able to realise and understand the criminal consequences which would follow therefrom. These capacities are both necessary and sufficient for holding a person responsible for their criminal conduct, because they are all that is required in principle for the law to fulfil its operant purpose of discouraging and preventing criminal behaviour. That is to say that any person possessing all three of these capacities, *in principle*, is fully capable of understanding the criminal nature of their behaviour and acting so as not to be in breach of a legal proscription.

The present theory of responsibility is also entirely consequentialist both in its foundations and implications. The first part of the thesis described the various mechanisms which feed into each component of a decision, mechanisms which can and do operate automatically. In simple terms, faced with a certain problem or decision to be resolved, these mechanisms represent different decision outcomes, recruit evidence from a variety of internal and external sources in support of these competing outcomes until one achieves supremacy, and then feed that decision outcome forward into the preparation and execution of the relevant motor actions to bring the decision into fruition.

Of course, this over-simplification betrays an unparalleled complexity to the brain's operations, not least when conducting high-level cognitive functions such as executive decision-making. For example, different networks are processing the *what, how, when, whether* and *why* components of each decision; each activated neuronal network representing a decision outcome is recruiting evidence from internal and external stimuli; the manner in which all of these stimuli are processed is equally moderated by countless additional factors, the operation of cognitive biases, the presence of certain beliefs or attitudes, the amount of energy one possesses in the particular moment of deciding, *etc.*

Nevertheless, the argument supported in the first part of this thesis is that decision-making, fantastically complex as it may be, is ultimately a deterministic process.

If *mechanisms* – a most deliberate choice of word – are responsible for processing stimuli and producing decision outcomes, it may also be relevant to consider from where such "stimuli" arise that are relevant to those decisions. An unquestionably significant contributor of such internal stimuli is memory, which provides a recollection of how we have dealt with similar problems and decisions in the past and is inherently built upon our past experiences. Similarly, our personality is a significant moderator of decision-making – people with more impulsive personalities inevitably make different decisions to those who are more deliberative. Indeed, all those internal factors which either provide the stimuli for decision outcomes or moderate how such stimuli are processed can ultimately be reduced to the culmination of our biology – genetics, epigenetics, homeostasis, psychology, personality, *etc.* – and our experiences up to the time of any given decision. Equally, the external stimuli which contribute to a decision consist of those things we have experienced or are experiencing; those things which have been, or are being, perceived and experienced in the past and in the moment.

All this is to say that, again, howsoever complex it is, human decision-making is ultimately deterministic. The outcome of any and all decisions is the product of prior causes, those causes consisting of each individual's unique biology (as established at birth in genetics and epigenetics), combined with the culmination of each individual's unique personal experiences to date. These two factors have both produced the unique decision-making architecture of every individual's brain, and ultimately also provide the stimuli that this architecture processes in order to result in a decision. As such, some form of decision not only rests as the cause of the resultant behaviour enacted to bring about that decision, but every such decision is equally an effect determined by a succession of prior causes. Thus, every decision is caused; every criminal decision (or state of *mens rea*) is equally caused; and so, necessarily, every non-criminal decision is caused (and theoretically causable) also.

## *12.1. Arguments Against Retributivism*

Retributivism encompasses a number of claims, but may be broadly described as the theory that punishment is a good in and of itself, independent of any corollary benefits such as preventing future crime, and that the law should punish those (and only those) who commit moral wrongs, proportionately to the wrong that is committed.[3] At the root of retributivist theories are the dual concepts of moral desert and proportionality. Punishment is reserved for those who are deserving thereof on account of having committed some moral wrong, both because the thing that they have done is wicked or immoral, and because they have *freely* chosen to do it. Further, punishment must be distributed in proportion to the offence that is being punished. This means that the innocent may never be punished for some greater good; however, proportionality may also be interpreted to restrict any allowance for mercy or clemency or, indeed, for harsher punishment of the guilty in pursuit of such further aims as general deterrence.[4]

There are many forms of retributive theories that have been developed, most encompassing the aforementioned broad descriptions whilst emphasising certain further aspects or arguments.[5] For example, substantially desert-based formations emphasise the simple fact that punishment is exacted because it is deserved, this being sufficient alone to justify the institution of punishment.[6] Communication-based formations of retributivism refer eponymously to the communicative nature of punishment and the role it plays in expressing society's condemnation of certain conduct.[7] And fair-play formations of retributivist theories focus on the self-restraint generally practised by people in society and the unfair advantages obtained by those who breach the rights of others.[8] In all instances, however, retributivism is a deontological theory because it

[3] See generally Michael S. Moore, 'Justifying retributivism' (1993) 27(1-2) *Israel Law Review* 15.
[4] See generally Thom Brooks, *Punishment* (Routledge 2012), 15 – 34.
[5] See Leo Zaibert, *Punishment and Retribution* (Ashgate Publishing 2006); John Leslie Mackie, 'Retributivism: A test case for ethical objectivity' in Feinberg J. and Coleman J. (eds.)., *Philosophy of Law* (6th ed. Wadsworth 2000); John Cottingham, 'Varieties of retribution' (1979) 29(116) *Philosophical Quarterly* 238.
[6] See R. Anthony Duff, *Punishment, Communication, and Community* (Oxford University Press 2003), 3 & 19 – 21.
[7] *Ibid*., 27 – 30 & Ch. 3; Andrew von Hirsch, *Censure and Sanctions* (Clarendon Press 1993), 41.
[8] See Martin P. Golding, *Philosophy of Law* (Englewood Cliffs 1975), 92.

identifies punishment as a moral right in and of itself, and is retrospective because it justifies punishment solely by reference to the wrongdoer's past actions.

Fassin notes the interesting etymological origins of "retribution" from the Latin *retribuere*, composed of *re-*, "in return", and *tribuere*, "to divide among tribes". Thus, the word originates from the notion of '"to give in exchange", "to pay in return", "to restitute what is owed" in the dual sense of recompense and punishing.'[9] Retribution in this sense was initially neutral, applying equally to giving reward for something earned or reprisal for something deserved, and subsequently took on opposing meanings in French and English. As Fassin explains, 'in the religious language of the time, *rétribution* refers to the fair salary rewarding merit, under the Calvinist influence, while retribution evokes the Last Judgment and God's Wrath, from a literal reading of the Bible.'[10] His argument is that retribution originates in the notions of reparation and repayment, evoking a meaning of exchange but *not* moral condemnation. It was not until the Renaissance that retribution became tied up with ideas of moral suffering in return for committing sin. It is this subsequent connection drawn between retribution and moral wrongdoing that is contested.

Kant is perhaps most closely associated with developing a moral theory of retributive punishment during the Enlightenment, although his ideas arguably remain substantially intertwined with Biblical themes. In one most famous passage, Kant writes:

> 'Even if a civil society were to be dissolved by the consent of all its members (*e.g.*, if a people inhabiting an island decided to separate and disperse throughout the world), the last murderer remaining in prison would first have to be executed, *so that each has done to him what his deeds deserve and blood guilt does not cling to the people for not having insisted upon this punishment*; for otherwise the people can be regarded as collaborators in this public violation of justice.'[11]

[9] Didier Fassin, *The Will to Punish* (Oxford University Press 2018), 47.
[10] *Ibid*.
[11] Immanuel Kant, *The Metaphysics of Morals* (Denis L. (ed.), Gregor M. (trns.), Cambridge University Press 2017), 116.

In another passage, Kant writes

> 'Punishment by a court… can never be inflicted merely as a means to promote some other good for the criminal himself or for civil society. It must always be inflicted upon him only *because he has committed a crime*.'[12]

Subsequent writers have provided their own interpretation of the underlying basis for Kant's retributivism. For example, Dolinko extrapolates perhaps the most purely deontological argument for Kant's retributivism, *i.e.,* 'the lawbreakers deserve punishment and that this, all by itself, constitutes a good or sufficient reason for the state to inflict punishment on them.'[13] From this perspective, the present thesis contends that a theory of punishment founded upon the notion of moral desert should follow a theory of responsibility which is not necessarily founded upon morality at all, whilst also contesting the degree of "freedom" that is generally required by retributivists in order to conclude that an individual is *morally deserving* of punishment.

Murphy considers that the concept of reciprocity underlies Kant's retributivism. In particular, in order for the law to operate justly, it is necessary to 'guarantee that those who disobey it will not gain an unfair advantage over those who do obey voluntarily' such that punishment, 'in its retribution, [] attempts to restore the proper balance between benefit and obedience.'[14] From this perspective, punishment is justified because it is necessary to redress the unfair advantage obtained by the criminal over the obedient. However, again, the present thesis contends that it is not so self-evident that criminals either generally begin from a position of relative equality, nor necessarily obtain some considerable advantage from their criminal conduct, sufficient for retributive punishment to be justified as a moral good in its own right.

---

[12] *Ibid*., 114 (original emphasis).
[13] David Dolinko, 'Some thoughts about retributivism' (1991) 101(3) *Ethics* 537, 542 – 543.
[14] Jeffrie G. Murphy, *Kant: The Philosophy of Right* (Mercer University Press 1994), 121; see alternatively Hans Saner, *Kant's Political Thought* (Ashton E. B. (trns.) University of Chicago Press 1973), 30 – 33.

### *12.1.1. Moral Wrongdoing*

The first contention that the present thesis holds against retributive theories of punishment is, primarily, the necessary connection that is drawn between moral wrongdoing as a basis of desert for punishment and, secondarily, the reliance that such moral desert places upon conceptions of free will.

Inherent to all varieties of retributive theory is the notion that it is the individual's moral wrongdoing which in and of itself justifies responding with punishment.[15] However, as has been highlighted in section 8.3 of this thesis, above, even current conceptions of legal responsibility in the common law tradition do not draw any *necessary* connection between criminal and immoral conduct, although the two inevitably often overlap. As should be readily apparent, this thesis does not introduce any additional link between law and morality, and / or punishment. Rather, the notion of individual responsibility is rooted in the operation of three mental capacities, whilst punishment follows the breach of a criminal rule in the presence of these capacities, justified by the teleological principle that legal rules require enforcement in order to retain their very character and function as legal rules. The particular point is that it is inconsistent that a system of legal responsibility should attribute responsibility without any necessary reliance upon moral wrongdoing, yet insist on moral wrongdoing as the justification for subsequent punishment.

From a more theoretical viewpoint, common objections to the retributivist link with moral wrongdoing invite the question, *which moral wrongdoing is relevant?* That is to say, how should it be decided which moral wrongs are worthy of punishment, for it surely cannot be said that all are so. This is not only a question of relativism – *i.e.,* in view of a plethora of possible moral theories denoting particular rights and wrongs, how should the "correct" theory be selected for the imposition of subsequent punishment. Even within any single moral theory, rights and wrongs are more often described as subsisting upon some sort of scale; it is not difficult to appreciate from any moral perspective that murder is worse than verbal assault. It follows that there must exist those moral wrongs which fall below a

[15] Michael S. Moore, *Placing Blame: A Theory of the Criminal Law* (Oxford University Press 2010), 91; see also David O. Brink, *Fair Opportunity and Responsibility* (Oxford University Press 2021), Ch. 6.

certain threshold for punishment, in which case it may be asked how indeed to select amongst those wrongs within a single moral theory that are deserving of punishment.

This question becomes all the less abstract in consideration of legal responsibility. If the law criminalises irrespective of moral wrongdoing whilst punishment can only be distributed on the basis of the same, a lacuna inevitably emerges with regards to offences which might be regarded as consisting of little or no moral wrongdoing *per se*, not least in relation to more administrative or regulatory offences. For example, it is not easy to argue that any great moral wrongdoing occurs when somebody fails to pay their television licence or road tax aside from the somewhat more abstract issue of freeriding. In a similar vein, it is not entirely clear that somebody who breaks the speed limit whilst driving down a straight, clear and demonstrably deserted road commits any considerable moral wrongdoing to necessarily attract criminal punishment.

Taken to a greater extreme, it is not unarguable that a "Robin Hood" character who stole from the wealthy to redistribute to the poor could be generating some form of moral *good* overall, notwithstanding their breaking the law, in which case a strict adherence to retributivism might even *preclude* the possibility of punishment for certain offences. Furthermore, opponents of drug criminalisation will argue not only that the individual consumption of drugs struggles to amount to any significant moral wrongdoing, but that the criminalisation of drugs itself produces considerably greater moral harm for individuals and society at large. Going further still, it can be argued that actively breaking a certain law *in protest* of its immorality may not only be a morally right thing to do, but a necessary thing to do in order to force change, such as some of the civil disobedience that has historically been instrumental during various civil rights movements. All this is to say that the retributivist appeal to moral wrongdoing is far from straight-forward with regards to identifying precisely which conduct can justify punishment.

One response might be to suggest that it is the breaching of a criminal prohibition *per se* that constitutes the relevant moral wrongdoing to justify punishment – *i.e.,* that adherence to legal rules is itself a moral obligation rendering their breach a wrongdoing. However, this raises the issue of immoral laws: even countries such as the UK or US that have not

fallen under totalitarianism have nonetheless historically maintained a menagerie of laws which would be considered immoral by modern standards, from those permitting slavery and later racial segregation and discrimination, to those preventing women's enfranchisement or the criminalisation of homosexuality, *etc.* At the other end of the spectrum are any number of totalitarian regimes that have targeted minorities *through the law*, as extensively occurred in Nazi Germany. The retributivist argument untenably maintains not only that somebody who *breaks* such an immoral law must be punished, but that it is *their punishment* which produces the moral good in the circumstances.

#### *12.1.2. Desert and Free Will*

Second to the issues highlighted with grounding a theory of punishment in moral wrongdoing, there equally arises the question of when a person may be considered sufficiently deserving of punishment once they have committed some recognised wrongdoing. This question is necessary, for example, to distinguish between accidents, reflexes, and other automatic or blameless behaviour, and intentional, volitional actions which result from considered choices. If two people pick up another's handbag not belonging to them and walk away, they have each *prima facie* committed the moral wrong of taking another's property. However, if one of those people has done so knowingly whilst the other only picked up the bag believing it to be their own, we immediately reassess each persons' relative desert for punishment and, although most legal systems will require a person to make civil restitution for their accidents, it is more unusual for mere accidents and mistakes to be criminally punished.

Retributivism generally responds to this distinction by attributing moral desert (or blame) on the basis of the individual's *free* choice – that is to say, a person deserves to be blamed for their moral wrongdoing *because* they have freely chosen to do wrong.[16] The "standard educated view" of moral wrongdoing and desert as it flows from Kant's moral insight follows:

[16] Thomas M. Scanlon, 'Giving desert its due' (2013) 16(2) *Philosophical Explorations* 101; John Martin Fischer, 'Desert and the justification of punishment' in Nadelhoffer T. A. (ed.), *The Future of Punishment* (Oxford University Press 2013).

> 'Our desert is determined by what we can control. On the Kantian principle that "*ought implies can*", we cannot fairly be blamed for what we could not have done differently because we could not control it. We cannot control the wind that carries our bullet or the child who happens to dart before our speeding car. *We can control whether we intend to kill someone with the bullet and whether we intend to drive fast despite the risk to children*.'[17]

Pereboom describes the sense of moral responsibility at issue as that containing the notion of "basic desert":

> 'For an agent to be morally responsible for an action in this sense is for it to be hers in such a way that she would deserve to be blamed if she understood that it was morally wrong, and she would deserve to be praised if she understood that it was morally exemplary.'[18]

Thus, Pereboom refers to "free will" as the 'strongest sort of control in action required' to achieve this sense of moral responsibility.[19] Caruso elucidates further the connection between moral desert and free choice:

> 'Understood this way, free will is a kind of power or ability an agent must possess in order to justify certain kinds of desert-based judgments, attitudes, or treatments – such as resentment, indignation, moral anger, and retributive punishment – in response to decisions or actions that the agent performed or failed to perform. These reactions would be justified on purely backward-looking grounds – that is what makes them *basic* – and would not appeal to consequentialist or forward-looking considerations…'[20]

[17] Moore (2010), 196 (emphasis added).
[18] Derk Pereboom, *Free Will, Agency, and Meaning in Life* (Oxford University Press 2014), 2.
[19] *Ibid*; citing Alfred R. Mele, *Free Will and Luck* (Oxford University Press 2006).
[20] Gregg D. Caruso, *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice* (Cambridge University Press 2021), 2.

Discussing the nature of the "will" and volition in particular, Moore places considerable emphasis on the requirement of consciousness in making a relevant free choice that may attract moral desert. He adopts a broad Lockean position that people attain personhood through possessing and exercising conscious experience; in particular, people can not only be consciously aware of live thoughts, beliefs, and desires *etc.* that are presently held in mind, but also have access to the pre-conscious from where such content may readily be summoned into conscious awareness, and lesser access still to the unconscious.[21] He argues further that one of the crucial functions served by volition is a resolving function – *i.e.,* faced with ordinary kinds of conflicting decisions actively requiring a response, volition form part of the hierarchy of intentions that resolve the question of "*what to do now*". In order to serve this function, Moore argues that volitions 'must be responsive to all (or at least a fair sample) of what one desires, believes, and intends' which is only available through consciousness; states of altered or disassociated consciousness 'seem to break the unity of consciousness that allows volitions to be formed that are responsive to all of one's desires, beliefs, and intentions, and not just responsive to a small subset.'[22]

The present thesis is certainly incompatibilist with regards to the free will-determinism debate; the thesis began with the underlying assumption precluding the existence of metaphysical free will within a deterministic universe, whilst free will has been shown not to be a requisite component of legal responsibility in any event in chapter eight of the present thesis. Equally, it has been argued that legal responsibility does not rest upon the immorality or otherwise of an agent's actions *per se*, but in their deciding and subsequently carrying out a course of conduct whilst in possession of the three capacities of being responsive to reason, possessing ordinary self-control, and appreciating the nature and consequences of their actions. Thus, the justificatory factors underlying retributivism – *i.e.,* moral wrongdoing and desert based on free will – are neither required in the current approach of the law to legal responsibility, nor supported by the evidence and arguments considered throughout the present thesis.

[21] Michael S. Moore, *Act and Crime: The Philosophy of Action and its Implications for Criminal Law* (Oxford University Press 2010), 151 – 155.
[22] *Ibid*., 258; see also Michael S. Moore and Heidi M. Hurd, 'Punishing the awkward, the stupid, the weak, and the selfish: The culpability of negligence' (2011) 5(2) *Criminal Law and Philosophy* 147.

Besides the question of free will, retributivism emphasises the conscious control of agents in decision-making. Conversely, the evidence considered in chapter six of the thesis reveals how the brain's control mechanisms can and do operate unconsciously, and the subsequently developed thesis does not rest legal responsibility upon any requirement for conscious control over decision outcomes. Indeed, beyond the rebuttable presumption of ordinary self-control over bodily actions contained within the idea of volition, no further specific control over the causes of choices is necessarily required by the present thesis. This "ordinary" capacity for self-control simply states that people generally possess the capacity to control their bodily actions to conform with their intentions. However, this description of control does *not* necessarily entail that people equally control the content of their will; to say that we can control bodily actions to conform with intentions does not equate to controlling what our intentions actually are in the first place.

It is further submitted that views regarding the necessary involvement of consciousness in volition and choice are not supported by a significant body of the evidence considered in Part One of this thesis. Evidence was considered in chapter three demonstrating how the brain can adopt and pursue goals unconsciously; in chapter four showing how the brain automatically produces plans to enact different decision options under consideration; in chapter five revealing how the unconscious brain reaches decisions of which the conscious mind later becomes aware; and in chapter six showing how the final choice of whether to enact or veto a particular decision can equally operate outside of consciousness. Where retributive theories justify punishment on the basis of people's *free* and *conscious* choices attracting moral desert, therefore, the significant body of evidence considered in the present thesis severely undermines the necessary involvement of consciousness in decision-making and action, and describes fundamentally deterministic (as opposed to metaphysically free) mechanisms by which brains make decisions.[23]

[23] See further Caruso (2021); Gregg D. Caruso and Stephen G. Morris, 'Compatibilism and retributivist desert moral responsibility: On what is of central philosophical and practical importance' (2017) 82(4) *Erkenntnis* 837; Elizabeth Bennett, 'Neuroscience and criminal law; Have we been getting it wrong for centuries and where do we go from here?' (2016) 85(2) *Fordham Law Review* 437; Derk Pereboom, 'Free will skepticism and criminal punishment' in Nadelhoffer T. A. (ed.), *The Future of Punishment* (Oxford University Press 2013); Joshua Greene and Jonathan Cohen, 'For the law, neuroscience changes nothing and everything' (2004) 359(1451) *Philosophical Transactions of the Royal Society: Biological Sciences* 1775;

## 12.2. Consequentialist Theories of Punishment

The present thesis is itself inherently consequentialist in nature – it opens with an assumption of universal causal determinism; describes the mechanistic and often automatic processes involved in decision-making, from the selection of goals and adoption of intentions to the choice between action options for fulfilling those goals; and proposes a theory of responsibility that rests upon three capacities deemed necessary and sufficient in order for the brain to make decisions that comply with rules. What is more, that capacity-based theory of responsibility is itself consequentialist in various regards – it identifies mental capacities that can be developed and taught to some degree (as, indeed, they are with children); that have a direct impact upon an agent's ability to act in compliance with legal, moral or other rules; and which can be the target of interventions and improvement through measures in the health, education and criminal justice systems.

More succinctly, the thesis reflects that whilst all decisions and actions are ultimately the result of prior causes, so future decisions and actions are similarly caused. Where the three capacities of reasons-responsiveness, ordinary self-control and appreciation of the nature of conduct are themselves necessary for holding an individual responsible for their past criminal behaviour, they can equally be the targets of interventions to prevent future criminal behaviour. Unsurprisingly, therefore, consequentialist theories of punishment are readily subsumed within the underlying ethos and overall perspective of the thesis.

### 12.2.1. Incapacitation

Arguably one of the fundamental purposes of the entire criminal justice system is to secure the general peace and safety of society and, in so doing, the law identifies particular unacceptable conduct as being criminal and imposes a range of measures intended to prospectively prevent that conduct from occurring. Punishment may consequently be justified when necessary to incapacitate individuals in order to practically prevent them from committing certain prohibited acts.[24] A wider conception of *incapacitation* 'relates

Russell Christopher, 'Deterring retributivism: The injustice of "just" punishment' (2002) 96(3) *Northwestern University Law Review* 843.

[24] Thomas J. Miles and Jens Ludwig, 'The silence of the Lambdas: Deterring incapacitation research' (2007) 23(4) *Journal of Quantitative Criminology* 287, 290.

to all sanctions and interventions aimed to impede, restrict or make impossible certain actions, without necessarily being accompanied by measures that aim at other goals and effects, such as retribution, rehabilitation, restoration, *etc.*'[25] The death penalty is the ultimate incapacitating punishment, as it prevents the offender from doing anything ever again, criminal or otherwise. Similarly obvious but less drastic, incarceration restricts all manner of liberties including where a person is physically confined, what they do and where they go each day, and who they meet and communicate with outside of prison. Incarceration, again, incapacitates a person from being able to commit further criminal offences (or anything else) amongst the general public.

Punishments for the purpose of incapacitation can also be less harsh and blunt, and more targeted and nuanced. In the UK, for example, Control Orders and Anti-Social Behaviour Orders can incapacitate offenders by requiring them to remain in their home between specified hours, or to register at their local police station at certain times of the day, thereby controlling their wider movements and general location. These orders may also be used to restrict a person's computer or internet usage and use of the telephone, thus placing restraints on both a person's movements and communications.[26] Similarly, electronic monitoring can be used to both track a person's movements and restrict them to certain places at certain times.[27] Further still, people may be banned from occupying certain offices or professions such as company director or doctor; this represents a yet more nuanced form of incapacitation to prevent people from committing offences in particular roles where they may have previously offended.[28]

[25] Marijke Malsch and Marius Duker, 'Introduction' in Malsch M. and Duker M.(eds), *Incapacitation: Trends and New Perspectives* (Routledge 2016), 2.

[26] See further Andrew Ashworth, 'Criminal law, human rights and preventative justice' in McSherry B., Norrie A. and Bronitt S. (eds.), *Regulating Deviance: The Redirection of Criminalization and the Futures of Criminal Law* (Hart Publishing 2009); Lucia Zedner, 'Preventative justice or pre-punishment? The case of control orders' (2007) 60(1) *Current Legal Problems* 174.

[27] See further Peter H. van der Laan, 'Part-time incapacitation: Probation supervision and electronic monitoring' in Malsch M. and Duker M.(eds), *Incapacitation: Trends and New Perspectives* (Routledge 2016).

[28] See further Marijke Malsch, Wendy Alberts, Jan de Keijser and Hans Nijboer, 'Disqualification from a profession or an office: Nature and actual practice' in Malsch M. and Duker M.(eds), *Incapacitation: Trends and New Perspectives* (Routledge 2016).

The incapacitation of dangerous offenders is readily justifiable on the basis of preventing future harm to others; for John Stuart Mill, this is the only purpose for which the State is justified in exercising its powers of incapacitation over individuals in society,[29] which would restrict imprisonment to the most serious of offences which interfere with another's personal safety and wellbeing, such as murder, injurious assaults and rape. More broadly, incapacitation may be justified to prevent persistent offenders from committing further crimes, and where necessary to secure people's attendance at other treatment or rehabilitation. Caruso and Pereboom develop a "public health quarantine" model of punishment drawing an analogy between dangerous criminals and people carrying dangerous infectious diseases.[30] Following this analogy, incapacitation is justified in societal self-defence – *i.e.,* when it is necessary to protect the wider society and individuals therewithin from harm caused by the offender.

There are, however, inherent factors which provide safeguards against, and limit the use of incapacitation. For example, considering that incarceration is the harshest punishment in the government's arsenal[31] carrying the greatest interference with the ordinary rights of an individual, it is arguable from proportionality that incarceration should be reserved for those cases where it is deemed most necessary – either to secure further medical or rehabilitative treatments or to protect society from harm likely further – and in response to the most egregiously harmful offences. Tonry contends that "moral panics" have contributed substantially greater to increased rates of incarceration than any corresponding increase in crime,[32] which warns against the *overuse* of incarceration as undermining its effectiveness towards overall crime prevention or reduction. Furthermore, incarceration is increasingly recognised as a relatively blunt tool in terms of preventing crime and, thus, pursuing the broader goal of the criminal justice system in protecting

[29] John Stuart Mill, *'On Liberty' and Other Writings* (Collini S. (ed.) Cambridge University Press 1989), 13; see also Ronald Dworkin, *Taking Rights Seriously* (Bloomsbury Academic 2013), 25.
[30] Gregg D. Caruso, 'Free will skepticism and criminal justice: The public health-quarantine model' in Nelkin D. K. and Pereboom D. (eds.), *The Oxford Handbook of Moral Responsibility* (Oxford University Press 2022); Derk Pereboom and Gregg D. Caruso, 'Hard-incompatibilist existentialism: Neuroscience, punishment, and meaning in life' in Caruso G. D. and Flanagan O. J. (eds.), *Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience* (Oxford University Press 2018); Gregg D. Caruso, *Public Health and Safety: The Social Determinants of Health and Criminal Behavior* (ResearcherLinks Books 2017).
[31] Accounting for the general prohibition of capital punishment across Europe.
[32] See Michael H. Tonry, *Thinking about Crime: Sense and Sensibility in American Penal Culture* (Oxford University Press 2004).

society. Some studies suggest that the prison social environment can be criminogenic resulting in increased rates of recidivism;[33] whist those studies more modest on the recidivist impact of prison nevertheless highlight that incarceration generally does not reduce criminal behaviour but does present a considerable economic cost to the public.[34]

Robust scientific evidence on the impact of incarceration on reoffending is surprising scarce, although a criminogenic effect of incarceration is widely theorised owing to the antisocial experiences and communities within prisons, and the stigma attached to convicts upon release. Nagin, Cullen and Jonson review several dozen relevant studies and conclude that incarceration has either a null or mildly criminogenic effect on reoffending and recidivism.[35] A more recent and illuminating review by Loeffler and Nagin suggested that most studies reveal little impact of incarceration on recidivism, whilst a smaller number of studies reveal mixed effects.[36]

Of particular note, however, is their conclusion that those studies revealing a positive impact of incarceration on recidivism concern settings where rehabilitation is emphasised in prison, whilst those studies suggesting a negative effect on recidivism concern circumstances of incarceration with little or no emphasis of rehabilitation. Interestingly, one study into the social interactions of juvenile offenders within prisons revealed a greater rate of recidivism for inmates who socialised with one another contrasted against those who isolated themselves, indicating towards the potential criminogenic effect of being exposed to and socialising with other offenders within the prison environment.[37]

---

[33] Gerald G. Gaes and Scott D. Camp, 'Unintended consequences: Experimental evidence for the criminogenic effect of prison security level placement on post-release recidivism' (2009) 5(2) *Journal of Experimental Criminology* 139; Francis T. Cullen, Cheryl Lero Jonson and Daniel S. Nagin, 'Prisons do not reduce recidivism: The high cost of ignoring science' (2011) 91(3Supp) *The Prison Journal* 48S.

[34] Paul Gendreua, Claire Goggin and Francis T. Cullen, 'The effects of prison sentences of recidivism' (Public Works and Government Services Canada Report, 1999-3).

[35] Daniel S. Nagin, Francis T. Cullen and Cheryl Lero Jonson, 'Imprisonment and reoffending' (2009) 38(1) *Crime and Justice* 115, 178.

[36] Charles E. Loeffler and Daniel S. Nagin, 'The impact of incarceration on recidivism' (2022) 5(1) *Annual Review of Criminology* 133.

[37] Michael Windzio, 'Is there a deterrent effect of pains of imprisonment? The impact of "social costs" of first incarceration on the hazard rate of recidivism' (2006) 8(3) *Punishment and Society* 341.

### *12.2.2.Deterrence*

A second theory of punishment with a clear consequentialist pedigree is that of deterrence, which claims that the purpose of punishment is to deter people from offending in the first place – *i.e.,* general deterrence – and to deter individual offenders from reoffending – *i.e.,* specific deterrence. Deterrence theories operate upon the basis that the unpleasurable aspects of punishment – the loss of rights, freedoms and liberties through incarceration or control orders, suffering financial or temporal penalties through fines and community service, and the public shame of bearing a criminal record – operate to persuade people against committing criminal offences. On this latter point, publicity can be an 'essential component of the general deterrence process', whilst increased sentencing for repeat offences can reflect specific deterrence.[38] When considering a particular course of action, deterrence theories assume that people take into consideration the chances of getting caught and the negative consequences thereof, and argue that the potential for and reality of punishment should outweigh the benefits of crime.[39]

Theories of deterrence can be attributed to the work of Enlightenment philosophers Cesare Beccaria[40] and Jeremy Bentham.[41] For Bentham – particularly noted for his work in the moral theory of utilitarianism and the philosophy of law – a theory of deterrence builds upon his underlying utilitarian thesis that people act to obtain pleasure and happiness, and to avoid pain and suffering. To the extent that criminal actions can bring about pleasure (or benefit), people would be more likely to commit criminal acts if there is no pain attached by way of punishment; ergo, punishment should aim to prevent recidivism in individuals and to deter people generally from committing criminal offences.[42] Again, however, Bentham identifies natural limitations on the use of the deterrence justification for punishment. For example, punishment must be appropriate (*i.e.,* proportionate) to the crime in order to have a deterrent effect; it must 'inflict more

[38] Stephen Brown, Finn-Aage Esbensen and Gilbert Geis, *Criminology: Explaining Crime and Its Context* (8th ed. Elsevier Inc 2013), 175 – 176.
[39] See further Thom Brooks, *Deterrence* (Ashgate Publishing 2014), Ch. 3.
[40] Cesare Beccaria, *On Crimes and Punishments and Other Writings* (Bellamy R. (ed.), Cambridge University Press 1995).
[41] Jeremy Bentham, *The Rationale of Punishment* (McHugh J. T. (ed.), *Prometheus Books* 2009).
[42] See further Gerben J. N. Bruinsma, 'Classical theory: The emergence of deterrence theory in the Age of Enlightenment' in Nagin D. S., Cullen F. T. and Jonson C. L. (eds.), *Deterrence, Choice, and Crime: Contemporary Perspectives* (Routledge 2018), 25.

pain than the profits of the crime' on the one hand, whilst remaining proportionate as 'equal punishment for unequal crimes often produces the commission of a worse crime' on the other hand.[43]

Bentham's theories have provided the philosophical underpinnings to modern economic theories of deterrence,[44] which view people as rational actors who, in pursuit of the greatest benefit to themselves, must calculate the respective utility of not committing a crime thereby foregoing any potential benefit; committing a crime and retaining the benefit by avoiding apprehension; and committing the crime and being apprehended and punished. What makes this utility calculation so complex is that the potential offender 'faces a *probability* of apprehension; in practice, an offender does not know if he or she will be apprehended.'[45] Equally, economic theories of deterrence have been applied to emphasise the limiting effect of a low probability of being caught on the deterrent effect of harsher punishments,[46] again introducing some requirement of proportionality in the application of deterrent punishments. Crucially, for the purposes of the present thesis, deterrent theories of punishment are readily supported for their consequentialist approach whereunder the focus remains on changing and diverting future criminal conduct.

Kessler and Levitt[47] offer an interesting and notable economic demonstration of the distinct incapacitation and deterrent effects of punishment in operation by looking at the changes in crime rate in California, US, immediately after the enactment of "Proposition 8" which enhanced prison sentences for the most serious offences. The incapacitating effect of Proposition 8 would only come into operation after standard prison terms had expired, meaning that any observed reductions in crime rates before standard prison terms

[43] *Ibid*.
[44] For example, see Gary S. Becker, 'Crime and punishment: An economic approach' (1968) 76(2) *Journal of Political Economy* 169; A. Mitchell Polinsky and Steven Shavell, 'The economic theory of public enforcement of law' (2000) 38(1) *Journal of Economic Literature* 45; David S. Lee and Justin McCrary, 'The deterrence effect of prison: Dynamic theory and evidence' (2017) 38 *Advances in Econometrics* 73.
[45] Aaron J. Chalfin and Sarah Tahamont, 'The economics of deterrence: A review of the theory and evidence' in Nagin D. S., Cullen F. T. and Jonson C. L. (eds.), *Deterrence, Choice, and Crime: Contemporary Perspectives* (Routledge 2018), 34.
[46] For example, see Eberhard Feess, Hannah Schildberg-Hörisch, Markus Schramm and Ansgar Wohlschlegel, 'The impact of fine size and uncertainty on punishment and deterrence: Theory and evidence from the laboratory' (2018) 149(1) *Journal of Economic Behavior and Organization* 58.
[47] Daniel Kessler and Steven D. Levitt, 'Using sentence enhancements to distinguish between deterrence and incapacitation' (1999) 42(S1) *Journal of Law and Economics* 343.

had run would distinctly reflect the deterrent effect of Proposition 8. The results found that the crime rate fell immediately upon enactment of the new law, representing the deterrent effect of punishment, whilst the crime rate continued to fall twice as much after three years, representing the additional incapacitating effect.[48] In particular, this research suggests that the *threat* of receiving an enhanced punishment, at least, can exert a general deterrent effect against those considering criminal activities.

One particular contention is whether or not harsher or longer sentences alone result in a greater deterrent effect, or whether deterrence occurs simply from the *possibility* of receiving a (harsher) criminal / custodial sentence in the first place. To this end, Mears, Cochran, Bales and Bhati analysed the relationship between sentence length and recidivism for more than 90,000 inmates released from prisons in Florida, US.[49] They found that recidivism increased in line with sentence length up to one year; then decreased as sentences increased up to two years; whilst sentence length appeared to have no effect on recidivism beyond two years. The results create a "U-shaped" relationship between sentence length and recidivism, with sentences of less than two years having a broadly criminogenic effect, whilst sentences above two years exerted little effect either way.[50]

Interpreted in terms of specific deterrence, sentences of less than two years displayed an anti-deterrent effect, whilst sentences above two years displayed a neutral effect. However, this pattern followed similarly for offenders under the age of 23 until sentence length exceeded 2.5 years, beyond which point longer sentences produced a greater rate of recidivism which, again in terms of specific deterrence, can be interpreted as an anti-deterrent effect.[51] This particular finding suggests that incarceration is significantly more

[48] See also Thomas J. Miceli, 'Deterrence and incapacitation models of criminal punishment: Can the twain meet?' in Harel A. and Hylton K N. (eds.), *Research Handbook on the Economics of Criminal Law* (Edward Elgar Publishing 2012).
[49] Daniel P. Mears, Joshua C. Cochran, William D. Bales and Avinash S. Bhati, 'Recidivism and time served in prison' (2016) 106(1) *Journal of Criminal Law and Criminology* 81.
[50] *Ibid.,* 117; see also Jacqueline Beard, Georgina Sturge, Maria Lalic and Sue Holland, 'General debate on the cost and effectiveness of sentences under 12 months and consequences for the prison population' (House of Commons Library, Debate pack CDP-2019-0063, March 2019), 4.
[51] *Ibid.,* 115.

criminogenic for younger offenders, raising the question of the appropriateness of custodial sentences for these offenders.

Pratt, Cullen, Blevins, Daigle and Madensen present a meta-analysis exploring the deterrent effects of four aspects of criminal justice, namely the certainty of being caught and punished, the severity of punishment, deterrence / sanction composites, and the effect of non-legal costs such as shame and social stigma.[52] They found that only the effects of certainty of punishment and non-legal sanctions were large enough to be considered as substantively deterrent whilst, even where the effects of severity of punishment and deterrence / sanction composites were statistically significant, they were 'too weak to be of substantive significance.'[53] These results reflect the findings of an earlier literature review by von Hirsch, Bottoms, Burney and Wikström concurring that the certainty of punishment exerts a broadly stronger deterrent effect than sentence severity.[54]

In light of the preceding research, this might be interpreted as representing a deterrent effect from the threat of punishment, but not from the mere enhancement of severity alone. Perhaps the starkest demonstration that longer sentences likely incur no significantly greater deterrent effect can be found in relation to the death penalty, undoubtedly the most severe punishment within the criminal justice system. Whilst it is highly likely that research in this area is significantly ideologically driven,[55] meta-analyses of a wide body of research are forthcoming which fail to find any significant deterrent effect of the death penalty, and even indicates towards an opposite anti-deterrent effect.[56]

[52] Travis C. Pratt, Francis T. Cullen, Kristie R. Blevins, Leah E. Daigle and Tamara D. Madensen, 'The empirical status of deterrence theory: A meta-analysis' in Cullen F. T., Wright J. P. and Blevins K. R. (eds.), *Taking Stock: The Status of Criminological Theory: Volume 15* (Routledge 2017).
[53] *Ibid*., 379.
[54] Andrew von Hirsch, Anthony E. Bottoms, Elizabeth Burney and Per-Olof Wikström, *Criminal Deterrence and Sentencing Severity: An Analysis of Recent Research* (Hart Publishing 1999); see also Aaron Chalfin and Justin McCrary, 'Criminal deterrence: A review of the literature' (2017) 55(1) *Journal of Economic Literature* 5.
[55] See Berit C. Gerritzen and Gebhard Kirchgässner, 'Facts or ideology: What determines the results of econometric estimates of the deterrence effect of death penalty? A meta-analysis' (2016) 4(6) *Open Journal of Social Sciences* 178.
[56] Stephen N. Oliphant, 'Estimating the effect of death penalty moratoriums on homicide rates using the synthetic control method' (2022) 0(0) [online] *Criminology and Public Policy* 1; Dieter Dölling, Horst Entorf,

### *12.2.3.Rehabilitation*

Rehabilitative theories of punishment are most aligned with the overall themes of the present thesis, claiming that 'if criminal conduct has been caused by certain factors and if those factors can be identified and appropriately remedied, then the offender can eventually be returned to society.'[57] Rehabilitation is inherently consequentialist – it recognises that just as past criminal behaviour has been caused, so future law-abiding behaviour may be caused through appropriate rehabilitative interventions. Rehabilitative theories can trace their origins to Plato, who proposed a tiered prison system which addressed respectively the seriousness of the crime committed and the rehabilitative potential of the offender. Whereas Plato recognised the importance and utility of deterrence as a theory of punishment, this was not adequate on its own, with rehabilitative approaches additionally required to justify the imposition of punishment upon others.[58]

Modern proponents of rehabilitative theories of punishment often emphasise the responsibilities of the State which run concurrent to its right to impose punishment upon criminal offenders in the first place. For example, Rotman argues in support of the rights of the offender to be rehabilitated and returned to society with the opportunity to be a productive citizen; thus, where the State assumes the right to punish offenders, it obtains a concurrent duty to rehabilitate them also.[59] Carlen emphasises that it is the *choice* to commit a criminal offence which entitles the State to punish offenders; however, she argues that the choices faced by offenders are often limited by circumstances of poverty and / or inequality, which are demonstrated to correlate with subsequent criminal behaviour.[60] Hudson argues that recognition must be given to the role that the State plays

Dieter Hermann and Thomas Rupp, 'Is deterrence effective? Results of a meta-analysis of punishment' (2009) 15(1-2) *European Journal on Criminal Policy and Research* 201.

[57] Raneta Lawson Mack, *A Layperson's Guide to Criminal Law* (Greenwood Press 1999), 10.

[58] Ronald Nold, Kelley Massingale and Omi Hodwitz, 'Justice in ancient Greece and Rome' in Hodwitz O. (ed.), *The Origins of Criminological Theory* (Routledge 2022); Mary Margaret Mackenzie, *Plato on Punishment* (University of California Press 1981), 213; Edward M. Peters, 'Prison before the prison: The ancient world and medieval worlds' in Morris N. and Rothman D. J. (eds.), *The Oxford History of the Prison: The Practice of Punishment in Western Society* (Oxford University Press 1995), 8.

[59] Edgardo Rotman, 'Beyond punishment' in Duff A. and Garland D. (eds.), *A Reader on Punishment* (Oxford University Press 1994).

[60] Pat Carlen, 'Crime, inequality and sentencing' in Duff A. and Garland D. (eds.), *A Reader on Punishment* (Oxford University Press 1994).

in contributing to certain causes of crime, which give rise to the responsibility of the State to take part in crime prevention, *inter alia*, by rehabilitating offenders.[61]

Notably, rehabilitation can potentially be justified even in circumstances where an individual is not held responsible for their actions, for example, where an individual has acted without self-control as a result of a significant mental illness and is, therefore, not guilty by reason of insanity. From this perspective, positivist schools of criminology recognise the offender as the "passive victim of external or internal forces", to which responding with rehabilitation is not only a punishment *per se*, but is an important meliorative tool for removing the causes of an individual's offending and returning them to being a productive member of society.[62] However, perhaps for this reason, a common objection to rehabilitation is that it treats a person's behaviour as the 'result of a "condition" to be cured rather than something for which they are responsible; and that this is to treat them as an object rather than a subject, as part of the natural world rather than as a free agent.'[63] This is the traditional Kantian objection to rehabilitation as a justification for punishment, because it views human action 'simply as a symptom of an underlying conditions that calls for a cure', and reduces the treatment of the individual merely as a means – (*i.e.,* as a means to altering offender behaviour) – as opposed to a means to an end in their own right as a responsible, deciding being.[64]

This argument against rehabilitation for treating the offender as a mere means is not persuasive. From a philosophical perspective, rehabilitation does not regard human decisions *merely* as a symptom of prior causes to be cured, but also gives recognition to the fact that future decisions can also be shaped or "caused" through rehabilitation. Far from being contrary to the dignity of the offender or treating them merely as a means, rehabilitation aims to realise the potential for an offender to be reformed into a moral and productive member of society, as opposed to simply a menace to be locked away or

[61] Barbara A. Hudson, *Understanding Justice: An Introduction to Ideas, Perspectives and Controversies in Modern Penal Theory* (2nd ed. Open University Press 2003).
[62] Gwen Robinson and Iain D. Crow, *Offender Rehabilitation: Theory, Research and Practice* (SAGE Publications 2009), 3 – 4.
[63] Christopher Bennett, 'Punishment and rehabilitation' in Ryberg J. and Corlett J. R. (ed.), *Punishment and Ethics: New Perspectives* (Palgrave Macmillan 2010), 56.
[64] *Ibid*., 56 – 57.

otherwise excluded from participation within society. The principal aim of punishment becomes addressing the underlying causes of criminality in order to return the individual offender to the beneficial position of living in society. Rehabilitation acknowledges the rights of the offender to access redress to the causes of their offending, and the responsibility of society in making best efforts to rehabilitate offenders, as opposed to simply locking them away.

A variation of this Kantian objection might ask whether rehabilitation can still be justified if it is imposed upon convicts against their will. In the first instance, it is highly doubtful whether many forms of rehabilitative treatment could be successful without the positive engagement of the offender[65] – for example, cognitive behavioural therapy would require some manner of engagement from the offender and a desire or willingness to actually reform their conduct. Nonetheless, more coercive means of rehabilitation can readily be imagined, such as enforced chemical castration as a treatment for sexual offenders. On the one hand, it is submitted that such forced rehabilitation could be justified where an offender has been found guilty of an offence and is thereby deemed responsible for their actions. The argument from consent suggests that by committing a criminal offence in full knowledge – or, at least, with full potential access to knowledge – of the consequences, including forced rehabilitative treatment, the convict has thereby consented to those consequences of their actions. A similar argument from a rights perspective suggests that the convict has forfeited their right to object to the coercive use of State power as a result of their offending; after all, convicts might readily object to all manner of coercive punishment including incarceration and monetary fines, but this alone does not mean that the State is not justified in continuing to punish the guilty.

On the other hand, it is readily arguable that rehabilitation cannot be imposed on defendants who have been found not to be responsible for their actions, having successfully argued a defence appealing to the diminution of one or more of the relevant crucial capacities for responsibility. Discussed further in sections 12.3.2 and 12.3.3,

[65] For example, see Karen K. Parhar, J. Stephen Wormith, Dena M. Derkzen and Adele M. Beauregard, 'Offender coercion in treatment: A meta-analysis of effectiveness' (2008) 35(9) *Criminal Justice and Behavior* 1109.

below, the only theories of "punishment" that are justifiable in the case of the non-responsible are those which make restitution to the victim,[66] rehabilitation, and incapacitation *as is strictly necessary for securing rehabilitation and for the protection of society from the defendant's likely commission of further offences*.

Punishment for special deterrence is irrational in principle when the offensive conduct was caused by a default in one of more of the crucial capacities, because those capacities are necessary in order for legal rules backed by the force of punishment to fulfil their teleological purpose of guiding behaviour.[67] Meanwhile, punishing the non-responsible defendant on the grounds of expressivism or general deterrence would treat the individual merely as a means for society's ends, raising the Kantian objection. For the same reason, *forcing* rehabilitation upon the non-responsible and non-consenting defendant also treats them merely as a means for society's ends. In the case where such rehabilitation is declined by the non-responsible defendant, incapacitation is available only as strictly necessary for the protection of society.[68] Whilst this still treats the non-responsible defendant merely as a means, the need for the law to safeguard the wider peace and safety of society at large unquestionably outweighs the Kantian objection; the only alternative would be to release into society all potentially dangerous psychiatric patients and permit anarchy to ensue.

From a more practical perspective, it is difficult to see how *any* theory of punishment would be acceptable if the traditional Kantian objection to rehabilitation is accepted. Incapacitation locks up offenders as a blunt tool for keeping society safe; deterrence uses the sentencing of individuals in particular cases as a means to deter others from offending generally; restitution imposes burdens upon the offender to make restoration to individual victims and the wider society; and the expressive function of punishment uses the

[66] Following the civil law of negligence, restitution is appropriate even in the absence of any finding of criminal responsibility on the part of the defendant because the rights of the innocent victim nonetheless warrant vindication and redress; the civil law is almost entirely disinterested in a defendant's "moral blameworthiness" or "state of mind" in circumstances where a recognised right of the innocent claimant has been breached.

[67] See further section 13.1.1, below.

[68] One caveat concerns where the non-responsible defendant *lacks capacity* to give or refuse consent to rehabilitation, in which case the courts may order such measures as are deemed to be in the defendant's best interest applying ordinary principles of mental health law.

sentencing of individuals as a means to making a statement of condemnation to society. Even retribution uses the offender merely as a means, treating their punishment as a good *per se.* Indeed, it seems *only* rehabilitation recognises the offender as a means to an end in their own right; punishment is acknowledged as a necessary evil for the greater safety of society and protection of law and order, justified by the rehabilitative aim of enabling the offender to act again as a means to their own end within society. If the Kantian objection to rehabilitation is accepted, it is hard to justify how any other theory of punishment is less objectionable on the same grounds, in which case no justification for punishment can be found.

The traditional Kantian objection to rehabilitative theories emphasises the offender as a means to their own end, whereas rehabilitation offends their basic human dignity in treating them merely as a means, and their actions merely as the effects of prior causes. For Kant, this is to deny what it is that makes humanity special, namely the capability to make free decisions; the Kantian objection therefore falls back on an assumption of free will that has been explicitly denied from the outset of the present thesis.[69] Alternatively, however, P. F. Strawson offers his own version of Kantianism which accepts the deterministic causal influences resulting in human choices, but argues that this does not 'exhaust our interest in human behaviour' nor necessarily provides the appropriate explanations to 'guide our interactions with people.'[70] In Strawson's view, therefore, what makes human choices and behaviour important is not that they are metaphysically free from prior causes, but that choice and behaviour gives rise to inevitable reactive attitude in others, and underlies the range of quintessential relationships that humans can have between one another which render them as the 'subjects of certain demands or normative expectations.'[71] Korsgaard elaborates the view more thoroughly:

> 'To hold someone responsible is to regard her as a person – that is to say, as a free and equal person, capable of acting both rationally and morally. It is therefore to regard her a someone with whom you can enter the kind of

[69] Bennett (2010), 57.
[70] *Ibid*; citing P. F. Strawson, 'Freedom and resentment' in Stawson P. F. (ed.), *Freedom and Resentment and Other Essays* (Routledge 2008).
[71] *Ibid*.

> relation that is possible only among free and equal rational people: a relation of reciprocity. When you hold someone responsible you are prepared to exchange lawless activity for reciprocity in some or all of its forms. You are prepared to accept promises, offer confidences, exchange vows, cooperate on a project, enter a social contract, have a conversation, make love, be friends, or get married. You are willing to deal with her on the basis of the expectation that each of you will act from a certain view of the other: that you each have your responses which are to be respected, and your ends which are to be valued. Abandoning the state of nature and so relinquishing force and guile, you are ready to share, to trust, and generally speaking to risk your happiness or success on the hope that she will turn out to be human.'[72]

Of all the theories of punishment considered, rehabilitation is arguably amongst the most consequentialist in nature, being aimed fundamentally at reforming an offender's future behaviour to ensure its compliance with the law, and thus enable the offender's subsequent *safe* return to society. Much research has therefore focused on identifying the "best" or most effective types of rehabilitative programs.[73] Lipsey, Landenberger and Chapman identify a number of consistent themes within this research: more effective programs target either criminal behaviour directly (such as through contingency management programs) or specific proximal causes of criminal behaviour (such as cognitive behavioural therapy ('CBT')). More effective programs use structured regimens as a primary component, reflected in the 'greater effectiveness of behavioral and skill-building programs for reducing recidivism'; multimodal programs are typically better than single treatment strategies; and relatively higher doses of rehabilitative treatment tend to be more effective, with the higher end of such treatment lasting greater than 25 weeks with 5 to 10 contact hours of treatment delivered per week.[74] In addition to

[72] Christine M. Korsgaard, *Creating the Kingdom of Ends* (Cambridge University Press 1996), 189 – 190.
[73] See Mark W. Lipsey, Nana A. Landenberger and Sandra J. Wilson, 'Effects of cognitive-behavioral programs for criminal offenders' (Campbell Systematic Reviews No. 6, 2007); Nana A. Landenberger and Mark W. Lipsey, 'The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment' (2005) 1(4) *Journal of Experimental Criminology* 451; Mark A. Lipsey, Nana A. Landenberger and Gabrielle L. Chapman, 'Rehabilitation: An assessment of theory and research' in Sumner C. (ed.), *The Blackwell Companion to Criminology* (Blackwell Publishing 2004).
[74] Lipsey, Landenberger and Chapman (2004), 219 – 220.

contingency management programs and CBT, further modes of rehabilitation may include other forms of psychotherapy, medical interventions such as sterilisation, treatment for addictions and enrolment in alcohol treatment programs, *etc.*

Concerning the overall effectiveness of rehabilitation as a form of punishment, hundreds (if not thousands) of individual studies have been conducted around the world investigating the efficacy of different rehabilitation methods and programs on rates of recidivism, with more than 40 further meta-analysis conducted.[75] Taking some examples, the "Reasoning and Rehabilitation" ('R&R') project[76] – developed in Canada and popularly exported around the world – was devised to provide offenders with training in a number of cognitive skills essential for pro-social adjustment, responding to the high incidence of deficiencies in such skills amongst offenders. One meta-analysis of research from four countries by Tong and Farrington found that the R&R program was successful in reducing recidivism in Canada, the US and the UK, within both community and institutional settings, and in relation to both low- and high-risk offenders.[77] The overall reduction in recidivism across all studies reached a significant 14% in comparison to controls; this increased to a 21% reduction in recidivism when the program was delivered in a community setting.[78]

An earlier meta-analysis by Pearson, Lipton, Cleland and Yee reported a 26% decrease in rates of recidivism in subjects undergoing the R&R program;[79] whilst the single largest study consisting of a sample of 2,125 convicts who completed the program reported a reduction in recidivism rates of up to 52.5% amongst medium-need cases, and 57.8% amongst sexual offenders.[80] Concerning the efficacy of behavioural and cognitive-

[75] Paula Smith, Paul Gendreau and Kristin Swatz, 'Validating the principles of effective intervention: A systematic review of the contributions of meta-analysis in the field of corrections' (2009) 4(2) *Victims and Offenders* 148, 149.
[76] Robert R. Ross, Elizabeth A. Fabiano and Crystal D. Ewles, 'Reasoning and rehabilitation' (1988) 32(1) *International Journal of Offender Therapy and Comparative Criminology* 29.
[77] L. S. Joy Tong and David P. Farrington, 'How effective is the "Reasoning and Rehabilitation" programme in reducing reoffending? A meta-analysis of evaluations in four countries' (2006) 12(1) *Psychology, Crime and Law* 3, 18.
[78] *Ibid.*
[79] Frank S. Pearson, Douglas S. Lipton, Charles M. Cleland and Dorline S. Yee, 'The effects of behavioral/cognitive-behavioral programs on recidivism' (2002) 48(3) *Crime and Delinquency* 476.
[80] David Robinson, 'The impact of cognitive skills training on post-release recidivism among Canadian federal offenders' (Correctional Service Canada, Research report no. R-41, 1995), 8.

behavioural rehabilitation programs more generally, Pearson *et. al.* found cognitive-behavioural therapy ('CBT') to be more effective than purely behavioural programs, reducing recidivism by around 30% when contrasted against controls. A later meta-analysis by Wilson, Bouffard and Mackenzie placed the effectiveness of CBT at between 20% to 30% reduction in recidivism compared to controls.[81] Providing a review of the meta-analyses and discussing the state of the research overall, Lipsey and Cullen consider that the effects of rehabilitation treatment are 'consistently positive and relatively large', albeit allowing for significant variability in those effects according to the type of treatment considered, the quality of its implementation, and the nature of the offenders being treated.[82] Similarly reviewing the body of meta-analyses, Smith, Gendreau and Swartz note the remarkable consistency of replication within the research, concluding that 'it is clear that treatments adhering to the principles of effective intervention are effective in reducing offender recidivism.'[83]

#### *12.2.4.Restoration / Restitution*

Rehabilitation is closely associated with the notion of "restoration" in the sense of returning the offender to a previous condition of non-offending.[84] Relatedly, restoration in the present sense – perhaps better termed as restitution – refers to returning the victims of offending to a previous condition as if they had not suffered the offence, so far as is meaningfully possible through compensation and offender-victim reconciliation programs. Restitution might also be regarded in a broader societal sense within which offenders make restoration to society for the costs – both financial and societal – of their offending, such as through community service and payback programs.[85] Thus, restitution may involve paying compensation to victims and punitive damages / fines to society; engaging in programs with the direct victims of crime; and making restoration to society

[81] David B. Wilson, Leana Allen Bouffard and Doris L. Mackenzie, 'A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders' (2005) 32(2) *Criminal Justice and Behavior* 172.
[82] Mark W. Lipsey and Francis T. Cullen, 'Correctional rehabilitation: A review of systematic reviews' (2007) 3(1) *Annual Review of Law and Social Science* 297, 297.
[83] Smith, Gendreau and Swartz (2009), 163.
[84] Robinson and Crow (2009), 1 – 2; see Landenberger and Lipsey (2005), 451.
[85] Lode Walgrave, *Restorative Justice, Self-Interest and Responsible Citizenship* (Routledge 2012), 38 – 40.

as an indirect victim of crime, for example, through providing free labour to the community.

Restorative justice has been the 'dominant model of criminal justice throughout most of human history for all the world's peoples.'[86] As was discussed in section 8.3 of this thesis, above, the concept of *mens rea* in the medieval law of murder was closely tied to the notion of making restoration to the family of the deceased victim by paying the *wergeld*. Modern proponents of restorative justice 'emphasize the need to support both victims and offenders and see social relationships as a rehabilitative vehicle aimed at providing formal and informal social support and control for offenders.'[87] Thus, restorative justice views the social support and control of offenders '*as the means* to rehabilitation.'[88] In this regard, a 2005 meta-analysis of restorative justice practices by Latimer, Dowden and Muise concluded that the results of available studies 'provide notable support for the effectiveness of these programs in increasing offender / victim satisfaction and restitution compliance, and decreasing offender recidivism.'[89]

Like rehabilitation, restitution does not necessarily require than an individual has been found to be responsible for their actions. The majority of the civil law of torts do not demand proof of responsibility – *i.e.,* volition, intention or other *mens rea* – in order to require that a tortfeasor pays damages, simply relying instead on the fact that the tortfeasor has caused some unreasonable harm to another.[90] Equally, the tortious measure

[86] John Braithwaite, 'Restorative justice' in Tonry M. H. (ed.), *The Handbook of Crime and Punishment* (Oxford University Press 1998), 323.
[87] Cyndi Banks, *Criminal Justice Ethics: Theory and Practice* (5th ed. SAGE Publications 2020), 158; citing Gordon Bazemore and Michael Dooley, 'Restorative justice and the offender: The challenge of reintegration' in Bazemore D. and Schiff M. (eds.), *Restorative Community Justice: Repairing Harm and Transforming Communities* (Anderson Publishing 2001).
[88] *Ibid*. (emphasis added).
[89] Jeff Latimer, Craig Dowden and Danielle Muise, 'The effectiveness of restorative justice practices: A meta-analysis' (2005) 85(2) *The Prison Journal* 127, 142; see also James Bonta, Rebecca Jesseman, Tanya Rugge and Robert Cormier, 'Restorative justice and recidivism' in Sullivan D. and Tifft L. (eds.), *Handbook of Restorative Justice: A Global Perspective* (Routledge 2006); Heather Strang, Lawrence W. Sherman, Evan Maro-Wilson, Daniel Woods and Barak Ariel, 'Restorative justice conferencing (RJC) using face-to-face meetings of offenders and victims: Effects on offender recidivism and victim satisfaction. A systematic review' (2013) 12(1) *Campbell Systematic Reviews* 1.
[90] Kylie Burns, Arlie Loughnan, Mark Lunney and Sonya Willis, 'Australia: A land of plenty (of legislative regimes)' in Dyson M. (ed.), *Comparing Tort and Crime: Learning from Across and Within Legal Systems* (Cambridge University Press 2015), 386 – 387; Jenny Steele, *Tort Law: Text, Cases, and Materials* (4th ed. Oxford University Press 2017), 30 – 32.

of damages aims to return the innocent party to the position that they would otherwise be in had the tort not occurred, and remove from the tortfeasor any benefit accrued from their actions.[91] This is justified because, even absent of the tortfeasor's intention or recklessness, the claimant is an innocent person who is affected by their actions and who reasonably ought to have been in the tortfeasor's contemplation, such as their "neighbour" for the purposes of the tort of negligence.[92] Regardless of the tortfeasors' responsibility (or *mens rea*), therefore, it is right that an innocent victim of their tortious actions is compensated to counteract the effects of those actions, so far as compensation is able to do so. By equal measure, restitution to the victims of criminal offending does not require proof of responsibility of the offender, because it is right that restitution is made to the victim (so far as it is possible to do so), regardless of whether or not the offender is *legally* responsible for their actions in the sense of exhibiting some culpable *mens rea*.

Restitution can be understood in both retrospective and consequentialist terms. With regards to the former, restitution is clearly backwards-looking insofar as it aims to make restoration for what the offender has done in the past, for the suffering caused to the victim and the cost of the offending to the wider society. Restitution aims to make good for a previous wrong. With regard to the latter, however, restitution also attempts to compensate both the victim of crime and the broader society for their continued and future suffering. Losses caused by past criminal activities have lasting effects for both the direct victims of offending and also wider society, not only financial but also physical and emotional. Most obviously, the direct victims of crime may suffer the continued financial effects of property lost through theft, or the continued emotional effects of offences against their person.

However, society too suffers from criminal conduct. Again, the most obvious costs might be those associated with the investigation and apprehension of offenders, and the cost of criminal prosecution before the courts. Furthermore, for example, hate crimes can instil fear and distrust in other members of a targeted group who are not the direct victims of a

[91] Mark Lunney, Donal Nolan and Ken Oliphant, *Tort Law: Text and Materials* (6th ed. Oxford University Press 2017), 881 – 882.
[92] *Ibid*., 111 – 118; citing *Donoghue v Stevenson* [1932] AC 562.

particular offence.[93] Similarly, areas of higher crime can suffer detrimental financial effects such as on property values,[94] with knock-on effects for the wider local community. Further still, the recent kidnap and murder in the UK of Sarah Everard at the hands of a serving police officer, (who specifically abused his police powers in the process), sparked significant outrage across the country and contributed to fostering great distrust between the public (and women in particular) and police. So far as is possible, restitution attempts to provide some degree of recompense for the continued and future effects of these and similar losses, in which regard this theory of punishment adopts a consequentialist outlook. Programs centred around offender-victim reconciliation and compensatory damages represent clear means of making restitution to individual victims. Meanwhile, community service and payback schemes, and punitive damages (in particular for financial offences, and against companies and high-net-worth individuals), represent ready mechanisms by which offenders can make restitution to society more generally.

With regards to the latter, punitive damages have traditionally been conceptualised as a deterrent punishment, inflicting a direct economic cost upon offenders in an effort to dissuade them from pursuing the gains of certain prohibited (civil or criminal) conduct. However, the actual deterrent effect of punitive damages is questionable, not least when applied to corporate offenders and high-net-worth individuals who may calculate the potential for punitive damages into the anticipated costs of proceeding with that prohibited conduct.[95] One novel response to this issue pioneered in Scandinavia and exported to countries around the world is the application of "day fines" which are calculated with different multipliers representing the seriousness or severity of an offence in question, which is then multiplied by the defendant's daily income. This arguably creates a stronger link between the punitive element of a financial penalty and the

[93] Mark A. Walters, Jenny L. Paterson, Liz McDonnell and Rupert Brown, 'Group identity, empathy and shared suffering: Understanding the "community" impacts of anti-LGBT and Islamophobic hate crimes' (2019) 26(2) *International Review of Victimology* 143; James G. Bell and Barbara Perry, 'Outside looking in: The community impacts of anti-lesbian, gay, and bisexual hate crime' (2015) 62(1) *Journal of Homosexuality* 98.

[94] Nils Braakmann, 'The link between crime risk and property prices in England and Wales: Evidence from street-level data' (2016) 54(8) *Urban Studies* 1990; Steve Gibbons, 'The costs of urban property crime' (2004) 114(499) *The Economic Journal* F441.

[95] For example, see James Boyd and Daniel E. Ingberman, 'Do punitive damages promote deterrence?' (1999) 19(1) *International Review of Law and Economics* 47; Jill Wieber Lens, 'Justice Holmes's bad man and the depleted purposes of punitive damages' (2013) 101(4) *Kentucky Law Journal* 789.

offender's actual ability to pay, resulting in significantly larger fines against wealthier offenders, and inducing a relatively stronger deterrent element to any financial calculus that is conducted to determine if a particular offence is worth committing.[96]

Compensatory damages paid towards the direct victims of crime are justified in both cases where a defendant is found guilty (and therefore responsible) for their criminal acts, and also where a defendant is found to be not responsible on account of some positive defence appealing to a deficiency in their relevant capacities. As discussed above, this follows because the innocent victim of offensive conduct is nonetheless deserving of compensation applying normal principles of civil law, which does not depend upon proof of culpable *mens rea*. However, it is submitted that punitive damages should only be available where a defendant is found guilty, in which circumstances they are responsible for committing criminal conduct in full possession of the relevant capacities for responsibility. It is only in this case that the *punitive* element of damages can reasonably be expected to have any desired deterrent effect upon behaviour, as the defendant lacking the relevant capacities cannot themselves reasonably be expected to conform their behaviour in accordance with legal rules. Moreover, the finding of legal responsibilty justifies the position that a guilty defendant should make restitution to the wider society, whereas to apply the same to the non-responsible defendant raises the Kantian objection of unjustly treating them as a mere means for serving society's ends.

#### *12.2.5.Declaration / Expressivism*

A central function of the criminal law is to declare those acts which are prohibited as being criminal, and express society's intolerance of those acts by punishing those who are responsible for committing them. Punishment therefore becomes an expression of public disapproval – it is 'not the mere infliction of pain but a statement of denunciation.'[97] Punishment can have a declaratory effect through the size of fines or length of prison sentences that are handed down for different offences, with larger fines / sentences expressing a greater degree of intolerance and condemnation of given offences.

[96] See Elena Kantorowicz-Reznichenko, 'Day-fines: Should the rich pay more?' (2015) 11(3) *Review of Law and Economics* 481.
[97] Thom Brooks, *Punishment: A Critical Introduction* (2nd ed. Routledge 2021), 118.

Punishment may also be declaratory through the imposition of a criminal record which extends beyond a period of imprisonment or payment of a fine, and provides a lasting public statement of condemnation for criminal offending.

Expressivism can be found in the jurisprudence of judicial giants. Victorian judge Sir James Fitzjames Stephen writes:

> 'The sentence of the law is to the moral sentiments of the public in relation to any offence what a seal is to hot wax. It converts into a permanent final judgment what might otherwise be a transient sentiment… the infliction of punishment by law gives definite expression and solemn justification to the hatred which is excited by the commission of the offence.'[98]

Seminal jurist Lord Justice Denning comments that the 'ultimate justification of punishment is not that it is a deterrent, but that it is the emphatic denunciation by the community of a crime.'[99] The philosopher Joel Feinberg has developed these ideas further in regarding the symbolic significance of punishment as:

> '[A] conventional device for the expression of attitudes of resentment and indignation, and of judgments of disapproval and reprobation, either on the part of the punishing authority himself or of those "in whose name" the punishment is inflicted.'[100]

Like restoration / restitution, declaration / expressivism can be regarded as both retrospective and consequentialist. With regards to the former, punishment operates as an expression of disapproval of past crimes committed. Moreover, society's condemnation as expressed through punishment must be proportionate to the crime committed: it is intuitively repugnant to express the same low level of condemnation of a murder as that given to someone breaking the speed limit, just as it seems plainly unjust to treat the

[98] James Fitzjames Stephen, *A History of the Criminal Law of England – Vol. II* (Macmillan and Co. 1883), 81.
[99] E. Gowers, *Report of the Royal Commission on Capital Punishment* (Cmd 8932, 1953), per Lord Denning.
[100] Joel Feinberg, 'The expressive function of punishment' (1965) 49(3) *The Monist* 397, 400.

speeder with the same high level of condemnation as the murderer.[101] With regards to its latter consequentialist effects, however, expressivism also makes forward-looking statements regarding what behaviour will and will not be tolerated by a society, and to what degree. As sentences for different offences are varied by the legislature and judiciary over time, and as new offences are added to the statute books and old ones removed, both the fact and severity of punishment operate to express society's changing views concerning different behaviours, and how much or little those behaviours are considered unreasonable and will be tolerated in the future.

Wrongful behaviour exists on a spectrum: at one far end are breaches of mere rules of etiquette such as incorrectly holding cutlery or chewing food with an open mouth which, whilst may be considered as reflecting a person's education, upbringing or social class, are scarcely reflections of character or "wrongdoing" in any real sense. Next might come breaches of common hygiene such as failing to cover the mouth whilst sneezing or not washing hands; such actions perhaps elicit greater disgust than breaches of etiquette, and may be judged more harshly for their potential to spread illness or disease. Next again follow more serious breaches of moral principles such as lying to friends and family; these breaches are generally considered more wrongful and likely to incur some degree of judgment and informal punishment such as shunning. The law begins to take substantive effect at the next stage of civil wrongs, such as in relation to breach of contractual agreements and various tortious actions; and, finally, the criminal justice system operates to prevent and deter the most serious of negative behaviours such as relating to property offences, battery and other violence, sexual offences and murder, *etc.*

It is further argued that such a spectrum of wrongful conduct continues within each category; few people would contest that any theft is "just as bad" as murder, although both are readily criminal offences. Thus, expressivist punishment functions as a pronouncement of the relative unreasonableness and society's varying intolerance of different offences. In this regard, it is submitted that punishments ought to be proportionate to the offence committed; again, most people would find it morally objectionable for a mother who steals baby milk formula to be punished as harshly as a

[101] See further Brooks (2021), 119.

murderer or, equally, for somebody who commits murder to be punished as leniently as another who shoplifts. The expressivist component of punishment "speaks" to each of the convicted defendant, their victim, and society at large. For example, the mandatory imposition of a life sentence for murder in the UK expresses to the defendant that they are reasonably expected to exert their capacities to the fullest extent to avoid this most serious of offences; to the victim (or, more correctly, their surviving family) that society recognises the grave and irreparable wrong committed against them; and to society generally that killing other people is amongst the most unreasonable and intolerable of offences.

### *12.2.6.Concluding Remarks on Consequentialist Theories of Punishment*

Thus far, the present chapter has argued specifically against retributivist theories of punishment on the principal basis that they are *solely* retrospective and rest upon notions of free will, conscious control of decisions, and moral responsibility. The present thesis has argued more broadly that neither free will nor moral responsibility are requisite constituents of legal responsibility and, it is therefore submitted, equally ought not be relied upon as requirements for punishment. The remaining broad overview of theories of punishment conducted in the present chapter has served predominantly to demonstrate that the remaining non-retributive theories of punishment can be justified on consequentialist grounds and, therefore, are acceptable within the present capacity-based theory of responsibility. The present thesis assumed the non-existence of metaphysical free will at the outset, elaborated on the implications of the denial of free will and consequentialism in chapter eight, and elaborates further on these philosophical implications of the thesis in the following chapter thirteen, below.

A common objection to consequentialist theories of punishment in general is that they can be extended to justify the punishment of the innocent for the greater benefit (*i.e.,* safety, security, lawfulness *etc.*) of society, for example, by imprisoning an innocent group of people for a particular offence in order to give the impression to society that the guilty offenders have been apprehended, and thus to deter future criminality by others. This argument is rejected on two grounds. First, it is submitted that the concept of

punishment necessarily follows responsibility, such that inflicting "punishment" upon the innocent is by definition not punishment at all, but almost certainly an offence in its own right. Feinberg gives a broad definition of punishment as 'the infliction of hard treatment by an authority on a person *for his prior failing in some respect (usually an infraction of a rule or command)*.'[102] Where the latter half of this definition focuses on the "prior failing" – *i.e.,* criminal conduct – of the individual being punished, it indicates that punishment is something that necessarily attaches to a person's criminal actions or behaviour. Punishment is by definition administered *for something that has been done* which, in the context of law or morality, is action that breaches a legal or moral rule respectively. No punishment of the innocent can therefore be justified because, by definition, this would not be "punishment" at all.

One method of demonstrating the necessary attachment of punishment to action is to draw a contrast against the concept of persecution. Where punishment is claimed to necessarily follow from something that a person has done (or failed to do when they were otherwise obliged to), persecution is more readily defined as a similar infliction of hard treatment, not in response to anything *done* by the persecuted individual(s) but, due to some characteristic of their person or membership of a particular group or community, often on religious, political, racial or sexual grounds.[103] Considering the example of the holocaust perpetrated during World War II, it is submitted that it would be incorrect to suggest that Jews, Romanies, Slavs, homosexuals and people with disabilities (amongst others) were *punished* for anything that they had done, whilst a far more accurate and precise use of language would be to say that they were *persecuted* on account of their status or membership of certain groups.

Equally, considering the subsequent Nuremburg trials, it would be quite inaccurate to suggest that members of the Nazi party were persecuted for their political beliefs, when they were charged, tried and punished for the things that they had done, *i.e.,* various war crimes and crimes against humanity. Thus, punishment of the innocent for some "greater

[102] Feinberg (1965), 397 (emphasis added).
[103] For example, see Ronald Christenson, 'The political theory of persecution: Augustine and Hobbes' (1968) 12(3) *Midwest Journal of Political Science* 419.

good" cannot be conceived as punishment at all, because punishment necessarily attaches to something that a person has done. This particular attribute of punishment and responsibility attaching to actions (or, more precisely, decisions to act) is explored in defended detail in the following chapter thirteen of the present thesis, below.

Second, it is submitted that the consequentialist theories of punishment considered each arrive with their own limiting factors and considerations. Most prominent amongst the various theories is the limiting concept of proportionality, which claims simply that punishment must be proportionate to the offence committed. From the perspective of deterrence, it is argued that disproportionate punishment diminishes its deterrent effect: why not rob a bank if it is punished the same as stealing some bread; and why not kill a rival if it is punished the same as simply punching them? From the perspective of incapacitation, this represents the greatest imposition upon the rights of the offender which, proportionately, should again be reserved for circumstances of greatest need, where incapacitation is required to secure the treatment or rehabilitation of the individual or the wider safety of society from the individual's likely reoffending.

From the perspective of rehabilitation, no further purpose is served by punishment that does not operate to address the causes of the offence and diminish the impact of those causes on potential future offending. In consideration of these and similar limiting arguments contained in each consequentialist theory of punishment, therefore, punishment of the innocent – for the reason that it may bring about certain beneficial consequences for the wider society – becomes unarguable. If punishment must be proportionate to the severity of the offence, for example, no offence committed at all must necessarily warrant no punishment at all. A more detailed and specific defence of the principle of proportionality is provided at section 12.3.3 of the thesis, below.

Continuing the present discussion, suppose that an unsolved violent offence has outraged a town to the point of revolt, in response to which the town sheriff arrests and executes an innocent person who is offered up as the guilty perpetrator. And suppose, for the sake of argument, that the sheriff's actions succeed in quelling the revolt which would otherwise undoubtedly have flared into civil unrest. Putting aside arguments concerning

the definition of punishment and the claimed necessary attachment of responsibility to decisions to act, could the sheriff's actions be justified in favour of the greater good of society? – after all, a principal function of the criminal justice system is to secure the peace and security of society which, in the absence of apprehending the actual offender, the sheriff has nevertheless achieved by alternative means. Even in such circumstances, it is submitted that punishment (or, more accurately, persecution) of the innocent for some "greater good" remains unjustifiable. It is recalled that the greater good concerns securing the peace and safety of society in relation to a revolt that has arisen in response to the unsolved violent offence, and the failure to apprehend the perpetrator.

It is submitted that punishment of the genuine offender is a sufficient means of securing the aforementioned greater good in the circumstances presented, whereas punishment of an innocent scapegoat is not sufficient. Allowing for the sake of argument that the punishment to be applied is some form of incapacitation – incarceration or even execution such that the punished individual will be incapable of committing any further offences – punishment of the genuine offender will readily meet the sufficiency condition. Whilst the town revolt will be quelled on the one hand, the genuine offender will be physically incapable of committing any further offences, such that it is impossible that *their* future actions could cause any recurrence of the revolt.

The same cannot be said in the case of punishing a scapegoat; the present town revolt may be quelled for a time, but the genuine offender remains at large with every capability of committing further offences. Not only is the recurrence of the relevant violent offending more likely (owing to the simple fact that the offender remains at large), but the town's revolt is similarly liable to reoccur in response to further offending. What is more, the argument is readily made that additional outrage may be caused in such circumstances when the town realises not only that the genuine offender remains at large, *but that an innocent member of their community has been unjustly punished in his stead*. Punishment of the innocent, therefore, is not sufficient in the circumstances presented to *ensure* or *guarantee* the peace and security of the society.

## *12.3. Verdicts*

There are three substantive verdicts generally available at the conclusion of a criminal trial in England and Wales;[104] a "guilty" verdict where the prosecution has successfully established the defendant's guilt beyond reasonable doubt, and a "not guilty" verdict where the prosecution has failed to prove their case to the requisite standard, or where the defendant has successfully pleaded a complete defence to the alleged charge. The third verdict is the special verdict of "not guilty by reason of insanity" which *must* be returned by the jury whenever they are satisfied that the relevant insanity defence has been successfully made out.

### *12.3.1. Reforming the Verdict of Not Guilty by Reason of Insanity*

The verdict of not guilty by reason of insanity used to require that a defendant be detained indefinitely within a psychiatric hospital, and still does in relation to murder. Otherwise, the judge is granted a wide discretion in sentencing, including hospital orders, restriction orders, supervision orders and an order for absolute discharge.[105] As Loughnan explains:

> 'The longstanding and intimate connection between the insanity doctrine and the special verdict has been explained as the result of a policy concern with marking out those defendants who are to be subject to the special coercive powers of the State from those who are either to be acquitted or convicted through the normal processes of the criminal law.'[106]

One particularly strong argument in favour reforming the current special verdict flows from the position that courts otherwise find themselves in. Wilson explains how for reasons largely of social defence, English courts have 'often strained to implement the

[104] An additional verdict of "not proven" is available in Scotland, which provides the same legal outcome as a not guilty verdict but may be delivered by the jury when they are unconvinced of the defendant's complete innocence.

[105] Criminal Procedure (Insanity) Act 1964, s. 5 (as amended by Domestic Violence, Crime and Victims Act 2004); see further Tony Storey, *Unlocking Criminal Law* (7th ed. Routledge 2020), 276.

[106] Arlie Loughnan, *Manifest Madness: Mental Incapacity in the Criminal Law* (Oxford University Press 2012), 166; citing Timothy H. Jones, 'Insanity, automatism, and the burden of proof on the accused' (1995) 111(Jul) *Law Quarterly Review* 475, 515; Eric Colvin, 'Exculpatory defences in criminal law' (1990) 10(3) *Oxford Journal of Legal Studies* 381, 392.

special verdict of not guilty by reason of insanity or even convict in cases where, although lacking definitional fault, the wrongdoer shows himself in need of treatment, supervision, or other corrective measures.'[107] In *R v Sullivan*,[108] for example,[109] the House of Lords held that the correct verdict was that of not guilty by reason of insanity in a case where the defendant had unconsciously and without subsequent recollection kicked another, due incontrovertibly to his suffering from an psychomotor epileptic seizure.

The Court expressed considerable sympathy for the defendant and was 'reluctant to attach the label of insanity to a sufferer from psychomotor epilepsy' of the kind suffered by the defendant, which consisted of a 'purely temporary and intermittent suspension of the mental faculties of reason, memory and understanding.'[110] Nonetheless, the Court was bound by primary legislation[111] and the label of "insanity" has accompanied any and all such cases attracting the special verdict of not guilty by reason of insanity. Indeed, there have been notable calls to reform not only the nomenclature of the insanity defence but also the way in which is applied in relation to certain conditions such as epilepsy.[112]

Considering the defences of insanity and automatism in particular, Wilson, Ebrahim, Fenwick and Marks[113] support the creation of a new form of special verdict. For reasons already identified in section 11.3.4, the insanity defence can peculiarly be both under- and over-inclusive with regards to the conditions that fall to be regarded as a disease of the mind. Meanwhile, courts have often needed to strain to apply the special verdict of not guilty by reason of insanity, or even resorted to applying a full conviction, to ensure that dangerous defendants fall under the necessary control of the courts, even if the individuals themselves substantively lacked any finding of fault. As the authors submit, on the one hand is the question of ensuring that only those deserving of punishment are

[107] William Wilson, 'How criminal defences work' in Reed A. and Bohlander M. (eds.), *General Defences in Criminal Law: Domestic and Comparative Perspectives* (Routledge 2016), 9 – 10.
[108] *R v Sullivan* [1984] 1 AC 156.
[109] See also *R v Burgess* [1991] 2 QB 92; *R v Quick* [1973] QB 910; *R v Hennessy* [1989] 1 WLR 287.
[110] *Sullivan* [1984], 173.
[111] Trial of Lunatics Act 1883, s. 2.
[112] For example, see R. D. Mackay and Markus Reuber, 'Epilepsy and the defence of insanity: Time for change?' (2007) (Oct) *Criminal Law Review* 782; R. D. Mackay and B. J. Mitchell, 'Sleepwalking, automatism and insanity' (2006) (Oct) *Criminal Law Review* 901.
[113] William Wilson, Irshaad Ebrahim, Peter Fenwick and Richard Marks, 'Violence, sleepwalking and the criminal law: Part 2: The legal aspects' (2005) (Aug) *Criminal Law Review* 614.

so punished whilst, on the other hand, there remains an inescapable need to protect society from people who may pose a threat, even if they are not at fault for so doing. They write:

> 'No one deserves punishment whose conduct was involuntary, unless such conduct was self-induced or the result of negligence. But desert also requires some recognition of the tragic narrative of violence in the absence of conscious awareness or volitional intent. This is a narrative which affects the perpetrator, who needs to expiate his sense of guilt, the victim, whose interest are unjustly set back, the victim's family who have to live with the consequences, and the wider society which needs the reassurance that the record is put straight and disorder replaced by order.'[114]

Approaching from a different perspective, Robinson considers the interaction between the criminal law's function in condemning certain conduct on the one hand, and the operation of excusatory defences on the other hand. Unlike justificatory defences which, in essence, hold that the defendant's conduct was justified and acceptable in the circumstances, excusatory defences essentially hold that that conduct was and remains unacceptable and condemnable, but that the individual defendant is excused on account of particular circumstances. Robinson argues that conviction in the case of excuses could nonetheless potentially be justified as a means of rehabilitating the offender and deterring both them and others from similar conduct in the future, whilst ostensibly harmful acts would continue to receive the sanction of the criminal law. However, whilst the criminal law condemns both the harmful conduct and the individual offender, this condemnatory function would be weakened if blameless defendants were punished who deserve to be excused from otherwise condemnable conduct. He submits, the 'only sound approach is to recognize excuse defences, but to minimize the danger of misperception of the acquittal by relying upon special verdicts – not guilty by reason of excuse – and assuring that the public understands their special message.'[115]

---

[114] *Ibid*., 623.

[115] Paul H. Robinson, 'Criminal law defences: A systematic analysis' (1982) 82(2) *Columbia Law Review* 199, 247.

### *12.3.2. The Verdict of Not Responsible*

The present thesis proposes the replacement and expansion of the current special verdict of not guilty by reason of insanity with a new general verdict of "not responsible". The not responsible verdict would be available in circumstances where: a) it is proven that the defendant committed the *actus reus* of a given offence; b) the defendant has successfully argued a defence which flows from causes that abrogated or overpowered any of the three crucial capacities for responsibility; and c) it is deemed necessary for the defendant to remain under the supervision of the court in order to address those aforementioned causes, for example, by compelling attendance at medical, psychiatric and / or rehabilitative treatment. In principle, the verdict would be available at the discretion of the jury; however, as is the case for other verdicts, the judge may direct the jury towards a verdict of not responsible in appropriate circumstances. Elucidating these three conditions, the first requires quite simply that the defendant actually did commit the act complained of, as must equally be proven for the verdict of guilty; a defendant who did not commit the *actus reus* of an offence is *de facto* and *de jure* innocent and, therefore, not guilty.

The second condition requires that the defendant has successfully argued a particular type of defence. The previous chapter eleven of this thesis has argued that each of the recognised defences may fairly be reasoned back to refer to one or more of the three capacities identified as crucial for responsibility. However, some of these defences are "circumstantial", in the sense that they arise from the specific circumstances surrounding the alleged offence as opposed to peculiarities of the individual defendant. Thus, defences of bare denial of *mens rea*, mistake, intoxication, self-defence, duress and necessity each arise in circumstances where it is recognised that the capacities of any individual would fairly and reasonably have been abrogated, overpowered or undermined. For example, self-defence is successfully argued in circumstances where *any* reasonable person is fairly expected to defend themselves or their family against violence; similarly, duress arises in relation to threats that are fairly accepted as overwhelming any reasonable person; and necessity in circumstances where any reasonable person would be forced to choose the lesser of two criminal options; *etc.*

Whereas these defences continue to relate to the crucial mental capacities, they are defences in which the circumstances surrounding the offence are what has impacted upon the capacities of the defendant, as opposed to the defendant suffering from some inherent deficiency in one or more of these capacities which itself led to their alleged offending. The circumstances – whether a person faces threats, events of necessity, makes mistakes, *etc.* – which underlie these defences are such that they ought to arise *and result in criminality* relatively infrequently for any normal law-abiding individual, such that the reoccurrence of those circumstances is not especially likely to be a cause of further offending in the future. Equally, those circumstances are typically such that the courts cannot reasonably undertake supervision. For a defendant who has successfully pleaded necessity, for example, the court cannot indefinitely supervise that defendant in case they should come up against such circumstances again giving rise to another criminal necessity in the future. It follows that a defendant who has successfully pleaded a bare denial of *mens rea*, mistake, intoxication, self-defence, duress or necessity should ordinarily proceed to a verdict of not guilty, (albeit the verdict of not responsible would in principle remain available at the court's discretion, discussed further below).

The point becomes clearer when considering those defences which would potentially warrant a verdict of not responsible; insanity, automatism, diminished responsibility, loss of control, and the proposed defence of addiction. Connecting these defences is how one or more of the crucial capacities of the defendant is undermined or overwhelmed by some aspect of their individual condition, rather than the particular circumstances of the specific alleged offence. Thus, each of these defences is significantly likely to be pleaded in circumstances where the defendant appeals to some particular condition, illness or abnormality which has a generally more persistent or permanent impact upon their capacities in order to give rise to the defence, and which likely contributed to the actual offending itself. For example, the defence of insanity refers specifically to the existence of an underlying medical condition such as schizophrenia; automatism may arise from other conditions such as a psychomotor epileptic seizure or hypoglycaemia; and the proposed defence of addiction eponymously relates to an underlying addiction disorder.

More to the point, a successful plea of insanity, automatism, diminished responsibility, loss of control, or addiction indicates conditions particular to the individual defendant that are liable, even likely, to result in their committing further offenses in the future. Moreover, these defences indicate conditions of the defendant that can potentially be treated and reformed, and can more reasonably be monitored and supervised by the courts to this end. It follows that the purpose of the not responsible verdict in response to the successful application of these defences is to nonetheless render the non-responsible defendant under the supervision of the court in order to address those underlying causes of their reduced mental capacities and consequent offensive behaviour.

The final condition of the not responsible verdict requires that it is deemed necessary for the defendant to remain under the supervision of the court. On the one hand, this condition reflects the fact that a given defendant has indeed been found *not to be responsible* for their otherwise criminal actions, in which case they might ordinarily expect an acquittal. On the other hand, this condition also reflects the overriding purpose of the criminal justice system in safeguarding the wider peace and security of society. In this regard, even people who are not responsible for their actions may sometimes need to be controlled in some manner for the greater protection of society. The not responsible verdict should therefore be given when the defendant has successfully relied on a defence which entails personal conditions that have so affected one or more of their crucial capacities such as not only to have caused their otherwise criminal behaviour, but which have likely potential contribute to future criminal actions if left unchecked or untreated.

#### *12.3.3. A Hierarchy of Verdicts and Proportionality in Punishment*

Having particular regard to the purpose of the not responsible verdict, discussed above, it becomes possible to draw a rough, three-tiered hierarchy of verdicts and punishment. At the bottom of the hierarchy is the verdict of not guilty, resulting in a total acquittal and no further action from the State on a particular issue or incident. At the top of the hierarchy is the verdict of guilty, following which the complete range of consequentialist punishments become available – incapacitation, deterrence (specific and general), rehabilitation, restitution and expressivism. More specifically, incapacitation can not only

be used as necessary to physically prevent further reoffending and thereby secure the safety of society, but in order to further other purposes of punishment such as securing attendance in rehabilitation or treatment programs, and to reflect the seriousness of the offending. Both specific and general deterrence against the guilty are justified – the former to provide good incentives to the individual offender and guide their future behaviour, and the latter to provide similar incentives to society more generally, and to dissuade others from criminal conduct. Further still, sentences following a guilty verdict (including fines, punitive damages and periods of community service) should be more or less onerous in proportion to the offence committed, expressing society's relative condemnation of different offences.

In the middle of the hierarchy rests the verdict of not responsible which, it is submitted, confines any subsequent "punishment" to the grounds of incapacitation *as strictly necessary in the circumstances*, rehabilitation and restitution. More specifically again, incapacitation in this case can only be justified when it is necessary to physically prevent the individual from likely committing further offences, and to secure that they attend any necessary treatments or rehabilitation ordered by the court. Neither specific nor general deterrence is justifiable; the former operates in principle by providing the individual with good reasons not to commit certain acts; however, the non-responsible defendant lacks one or more of the three capacities required for any such good reasons to take effect on their decision-making.

Meanwhile, to apply general deterrence would be to restrict the rights of the individual not to prevent *their* future offending or secure *their* rehabilitation, but as a signal to others not to offend. This raises the Kantian objection of treating the non-responsible individual merely as a means for the general deterrence of others and not as an end in their own right. Moreover, the signal given to society is objectional, being that harsher deterrent punishments will be administered regardless of an individual's responsibility for their actions. Expressivism similarly cannot be justified for the same reasons of treating the non-responsible individual merely as a means of expressing society's intolerance to others. Meanwhile, restitution remains available on civil law principles of making good to the innocent victims of offending, even if the offender was not responsible. However,

restitution to society by way of punitive damages or community service are again impermissible.

Aside for incapacitation *as is strictly necessary*, rehabilitation provides the quintessential purpose of punishment following a verdict of not responsible. As discussed in the previous section, this verdict exists *inter alia* because although the defendant's defence has succeeded, it has done so by appealing to causes that have undermined or overwhelmed their crucial capacities resulting in their otherwise criminal conduct, *and those same causes are liable to contribute to further criminal conduct if not addressed*. The principal purpose of the not responsible verdict is to render the defendant subject to the further supervision of the court such that these causes of their past criminal behaviour can be treated, rehabilitated or reformed in order to prevent future criminal behaviour, and so that the individual can be returned to society as swiftly but safely as possible. Thus, whereas the application of coercive State power against the responsible guilty defendant is justified across all of the consequentialist theories of punishment, the similar application of coercive power against the non-responsible individual can only be justified as is strictly necessary for the broader safety and security of society. This recognises that the non-responsible defendant is still being used as a means to society's end on the one hand whilst, as a matter of pure pragmatism, greater harm would likely ensue if such dangerous (albeit non-responsible) individuals were simply released back into society.

*

Reference has been made to the principle of proportionality throughout the present thesis and in this chapter on punishments in particular. Proportionality is one of the oldest principles in jurisprudence; proportionality as commutative justice – *lex talionis* (*i.e.,* "an eye for an eye") – is traced to the Code of Hammurabi, whilst proportionality as distributive justice finds its origins in Aristotle's *Nicomachean Ethics*.[116] Whereas the principle is often closely linked to retributive theories of justice – rejected in section 12.1

[116] Eric Engle, 'The general principle of proportionality and Aristotle' in Huppes-Cluysenaer L. and Coelho N. M. M. S. (eds.), *Aristotle and the Philosophy of Law: Theory, Practice and Justice* (Springer Science and Business Media Dordrecht 2013), 265; citing Aristotle, *Nicomachean Ethics* (Bartlett R. C. and Collins S. D. (trns.), University of Chicago Press 2011), Book 5.

of the present thesis – the retributive notion of proportionality is concerned with commutative justice.[117] Today, commutative justice is more properly the preserve of the civil law of contract and tort, where "justice" is achieved through the principle that 'no one should gain by another's loss' and by 'restoring the *status quo ante*.'[118] Conversely, distributive justice is more closely concerned with the criminal law wherein proportionality refers to the notion that punishment should "fit" the crime:

> 'Clearly, in the case of punishments, we "distribute" burdens of different gravity to people who harm others, just as in the distribution of rewards and prizes we distribute goods to people in proportion to what we regard as their desert. Punishment "fits" the crime not in the sense that it is equal to it but only in the sense that it remains an adequate proportion to other punishments for other offences.'[119]

Thus, the principle of proportionality can be preserved within the present theory of responsibility, notwithstanding that retributivism has been rejected. Just punishment within the criminal law is principally concerned not with *lex talionis,* restoring some manner of formal equality or *status quo ante* between an offender and their victim, but with the principle that 'State action must be a rational means to a permissible end which does not invade protected human rights unless strictly compelled by necessity.'[120] Indeed, it might readily be argued that the criminal justice system itself emerged in part as a rejection by civil society of *lex talionis* and the meting out of individualised "vigilante" justice between feuding private parties, to be replaced with the formal administration of institutionalised State power for the purposes of establishing a more civilised, fair, just and peaceful society.

Proportionality in criminal justice serves a number of moral, practical and even economic purposes, discussed in no particular order. First, continuing from the above discussion,

---

[117] Wojciech Sadurski, 'Social justice and legal justice' (1984) 3(3) *Law and Philosophy* 329, 331 & 334 - 335; Morris Ginsberg, *On Justice in Society* (Penguin 1965), 71 – 73.
[118] *Ibid.,* 335.
[119] *Ibid.,*
[120] Engle (2013), 265.

proportionality may be claimed as a principle of justice in its own right. The imposition of punishment by the State necessarily involves interfering with the fundamental rights of the convicted defendant. This remains the case whether it justified by some species of consent or contract theory, a voluntary relinquishment of rights by the convicted, the vindication of rights of the victim, or the broader protection of society in general. This further remains the case whichever theory of punishment is implemented – incapacitation, deterrence, rehabilitation, restitution or expressivism. Thus, applying the Aristotelean principle of proportionality as developed with regards to defensive force by Cicero,[121] Justinian,[122] Augustine[123] and, in particular, Aquinas,[124] any use of force must be necessary, exercised according to rules, and must not be excessive to the purpose for which it is applied.[125] The same argument is readily extended to States when the force of punishment is used in the enforcement of criminal laws against citizens.

A related moral argument in defence of proportionate punishment follows from the rule of law, another ancient legal principle[126] which traces its origins to Aristotle and ancient Greece[127] and was subsequently developed in Roman law by scholars such as Cicero.[128] The "modern" restatement in English law is provided by A. C. Dicey who describes the rule of law in its simplest form as the 'absolute supremacy or predominance of regular law as opposed to the influence of arbitrary power,' and 'equality before the law, or the equal subjection of all classes to the ordinary law of the land administered by the ordinary Law Courts.'[129] Lord Bingham disambiguates the rule of law into further sub-principles, proposing that it requires the law to be accessibly promulgated and operate in a predictable manner; that the law ought to protect fundamental human rights; that public

[121] Marcus Tullius Cicero, *The Republic and the Laws* (Rudd N. (trns.), Oxford University Press 1998), 69 - 70.
[122] Flavius Petrus Sabbatius Justinian, 'The *Lex Aquilia*' in Watson A. (trns.), *The Digest of Justinian: Volume I* (University of Pennsylvania Press 1985), 291.
[123] Augustine of Hippo, *The City of God* (Dods M. (trns.), Hendrickson Publishers 2009), Book XIX, Ch. 7.
[124] Thomas Aquinas, *Summa Theologica: Volume I – Part I* (Fathers of the English Dominican Province (trns.), Cosimo Inc. 2007), 458 – 491 (questions 90 – 97).
[125] Eric Engle, 'The history of the general principle of proportionality: An overview' (2012) 10(1) *Dartmouth Law Journal* 1, 4 – 5 (and footnotes 12 – 15).
[126] See Brian Z. Tamanaha, *On the Rule of Law* (Cambridge University Press 2004), Ch. 1.
[127] John Walter Jones, *The Law and Legal Theory of the Greeks: An Introduction* (Clarendon Press 1956), 90; Aristotle, *Politics* (Jowett B. (trns.), Dover Publications 2000), 139.
[128] Cicero (1998), 150.
[129] Albert V. Dicey, *The Law of the Constitution* (Allison J. W. F. (ed.), Oxford University Press 2013), 119.

power is exercised in accordance with the law (as opposed to being arbitrary); and that legal procedures are fair.[130]

A necessarily consequence of the rule of law, therefore, is contained in the maxim that "like cases are treated alike", which equally implies the opposite that different cases are treated differently. Just as it would offend the rule of law to imprison one murderer whilst permitting another to walk free when the circumstances of each case are otherwise equivalent in every meaningful way, so the rule of law is offended when the murderer and the petty thief are given exactly the same treatment in law despite the gulf of differences between these offences. More broadly, the demands of the rule of law in securing fundamental rights readily suggests that the State must be proportionate when interfering with those rights; the demands in treating people equally requires that the law balances competing rights claims proportionately when two different claims enter into conflict, including such conflict as arises between the rights of the individual and the rights of the State itself. Indeed, following a broad comparative review of constitutions around the world, Beatty goes so far as to assert that the principle of proportionality represents the "ultimate" expression of the rule of law.[131]

Addressing an eminently more practical argument in favour of the principle of proportionality, it is submitted that disproportionate punishment can exert an anti-deterrent effect and actually exacerbate the commission of further criminal offending. The logical argument follows that, where an offender faces proportionate punishments for offences of varying severity, it is rational for them to select the least serious offence which nonetheless achieves their particular goal in order to mitigate and reduce the severity of punishment in the event that they are caught and prosecuted. Conversely, if an equally lenient *or* severe punishment is applied regardless of the seriousness of their offending, there is no incentive to select the less over the more serious offence when they will be treated the same nonetheless, *especially if the more serious offence is more likely the fulfil their particular goal.* Suppose a gangster could intimidate and threaten their rivals at the risk of going to prison for a few years, or kill their rivals at the risk of going

[130] Thomas Henry Bingham, 'The rule of law' (2007) 66(1) *Cambridge Law Journal* 67, 71 – 79.
[131] David M. Beatty, *The Ultimate Rule of Law* (Oxford University Press 2004), Ch. 5.

to prison for life. If the punishment for either option is the same – whether that is "merely" imprisonment for a short term of years, or imprisonment for life – punishment itself offers no incentive to select the less severe option.

Some evidence for this argument can be drawn from the implementation of "three strikes" laws in the US; in essence, these laws impose a mandatory sentence of 25 years' incarceration or longer once somebody commits their *third* felony offence. In this respect, therefore, the punishment administered for a single third offence may readily be highly disproportionate to the seriousness of that offence, notwithstanding that the rule is justified by the commission of two previous offences.[132] Marvell and Moody investigated data derived from 50 US States over 29 years between 1970 to 1998, focusing on homicide offences in particular.[133] Their results found that the implementation of three-strikes laws was associated with a short-term increase in homicide rates of 10% to 12%, and a long-term increase of 23% to 29%; the implementation of each law across 24 States resulted in approximately 60 additional homicides in the short-term, translating into approximately 1,400 additional homicides across the 24 implementing states, and 1,200 lives "saved" across the 26 States which did not implement similar laws. The three-strikes laws were calculated to contribute to approximately 3,300 additional homicides per year in the 24 implementing States over the long-term.[134]

Comparable findings are reported by Sloan and Vieraitis exploring data from across 188 large US cities between 1980 and 1999.[135] Their results suggested a short-term increase in homicides of 13% to 14% in cities implementing the three-strikes laws, and a long-term increase of 16% to 24%, contrasted against cities without such laws.[136] Research by Chen focused specifically at data from California, US, as contrasted with other States

[132] In California, US, it is estimated that approximately 56% of offenders incarcerated under the three-strikes rule are convicted for less serious and / or non-violent offences; see California Legislative Analysts' Office, 'A primer: Three strikes – The impact after more than a decade' (*Legislative Analyst's Office,* October 2005) <https://lao.ca.gov/2005/3_strikes/3_strikes_102005.htm> accessed 19 October 2022.
[133] Thomas B. Marvell and Carlisle E. Moody, 'The lethal effects of three-strikes laws' (2001) 30(1) *Journal of Legal Studies* 89.
[134] *Ibid.*, 96.
[135] Tomislav Kovandzic, John J. Sloan and Lynne M. Vieraitis, 'Unintended consequences of politically popular sentencing policy: The homicide promoting effects of "three-strikes" in US cities 1980 – 1999' (2006) 1(3) *Criminology and Public Policy* 399.
[136] *Ibid.,* 409.

between 1986 and 2005, where the three-strikes law has received the largest implementation by a considerable degree over other States.[137] On the one hand, the results showed that rates for robbery fell 3% more quickly in California than comparison States, 1.8% more quickly for burglary, 1.1% for larceny-theft and 2% for motor vehicle theft.[138] However, offending rates for these offences were generally declining nationwide at the time whilst, curiously, such non-violent offences as burglary, larceny and vehicle theft were not eligible for enhanced sentencing in most States under the three-strikes laws in any event.[139] On the other hand, rates of murder increased 12.9% more rapidly in California than comparison States, consistent with both the magnitude and direction of effects found in previous studies.

Indeed, all three of the studies reported here hypothesise that the 'fear of a long mandatory sentence may motivate some criminals to attempt to eliminate witnesses or resist law enforcement officers.'[140] This hypothesis finds some support in a survey study by Schafer gathering responses from 604 juvenile offenders in California.[141] Most specifically, when subjects were asked whether, 'since I am going to prison for life if I get caught, I may as well kill any witness(es) because I have nothing to lose and I may go free if there is no one to testify', 54% of respondents answered 'yes.'[142] Thus, it is submitted that punishment disproportionate to the offence can indeed exhibit an anti-deterrent, criminological effect. In this regard, it is recalled that a link between the increased severity of punishment and subsequent deterrent effect was not readily forthcoming from research considered in section 12.2.2 of this thesis, above. However, the present argument suggests that the *disproportionality* of punishment can itself result in an *increased severity* of offending, providing practical support for the principle of proportionality in punishment generally.

[137] Elsa Y. Chen, 'Impacts of "three strikes and you're out" on crime trends in California and throughout the United States' (2008) 24(4) *Journal of Contemporary Criminal Justice* 345.
[138] *Ibid.,* 357.
[139] *Ibid*.
[140] *Ibid.,* 360.
[141] John R. Schafer, 'The deterrent effect of three strikes laws' (1999) 68(1) *FBI Law Enforcement Bulletin* 6.
[142] *Ibid.,* 8.

Finally, an obvious economic argument in favour of proportionate punishment may be made. State governments almost invariably operate under restricted financial resources, whilst almost all State finances arises from a combination taxation on citizens (including corporate citizens) and raising debt, which itself is repaid (with interest) through taxation. Meanwhile, it is trite that the criminal justice system is expensive to administer; in particular, the harshest punishment of incarceration also falls amongst the most expensive, whilst less onerous responses such as the imposition of financial penalties and community service and payback programs are invariably less expensive to administer.[143] To the extent that responses such as rehabilitation, and community service and payback programs also show more promising results on recidivism and reoffending rates, a cost-benefit to these less severe interventions over the alternative of incarceration is readily implied.[144] It is therefore not difficult to make the economic argument for proportionate punishment, with a view to both minimising overall expenditure on criminal justice to that which is necessary (thus easing pressure on government finances), and ensuring the most efficient allocation of that criminal justice expenditure towards those punishments which are most effective in proportion their relative cost.

[143] For example, see Kevin Marsh and Chris Fox, 'The benefit and cost of prison in the UK. The results of a model of lifetime re-offending' (2008) 4(4) *Journal of Experimental Criminology* 403.
[144] For example, see Jay Gormley, Melissa Hamilton and Ian Belton, 'The effectiveness of sentencing options on reoffending' (UK Sentencing Council 2022).

# 13. Philosophical Placement of the Present Thesis

'All theory is against the freedom of the will; all experience for it.'

- Samuel Johnson, 1778.[1]

The present thesis began with the fundamental assumption that the causal determinism of the universe is true, and that this truth precludes the possibility for metaphysical free will, defined according to the principle of alternative possibilities and the possibility for human decisions to be an "original," uncaused cause, or *causa sui* of events in the world. Conversely, current conceptions of legal responsibility – and *mens rea* within criminal liability in particular – are broadly constructed upon the opposite foundational assumption that human decisions are free. In this respect, proof of subjective mental states (*mens rea*) in combination with the presumption of volition attracts moral blame because the individual in such circumstances is taken to be the originating "author" of their own choice, and is deemed to have been capable of making a different choice in the circumstance. In light of the assumptions against free will adopted from the outset of this thesis, and the subsequent neuropsychological research describing largely automatic, mechanistic processes which result in decisions to act, the thesis has proposed replacing the current approach to *mens rea* with a capacity-based theory of legal responsibility.

This final substantive chapter of the thesis re-enters into the broader philosophical debate regarding free will and determinism. In particular, it is commonly asserted that the incompatibilist, "hard" determinist stance adopted by this thesis generally precludes not only the possibility for free will (which is herein agreed), but with it the possibility for moral responsibility for actions. This latter claim is clearly refuted insofar as it applies to the law; however, the present chapter goes further to refute this claim entirely. In so doing, it is submitted that the theory of *legal* responsibility here presented can readily be

[1] James Boswell, *The Life of Samuel Johnson LL. D.: Including a Journal of a Tour to the Hebrides* (Croker J. W. (ed.), George Dearborn & Co. 1833), 169.

generalised to cover *moral* responsibility also and, indeed, any decisions to act (legal, moral or otherwise) carried out by human beings. After generalising the capacity-based theory of responsibility, the present chapter engages with a number of leading discussions in the contemporary philosophical debate concerning free will, determinism and, in particular, responsibility for decisions and actions.

Four key propositions are defended: that the capacity-based theory of responsibility is justified by the fundamental teleology of any legal or moral rules concerning human actions and behaviour; that this teleology remains entirely defensible without any necessity to invoke concepts of metaphysical free will; that responsibility attaches to decisions to act and is not undermined by the fact that such decisions are causally determined; and, rather, that it is causal routes which significantly overpower or abrogate entirely any of the three discussed mental capacities which undermines responsibility and provides absolution for decisions to act taken in such circumstances. Together, these propositions will show how people can rationally be held responsible for their actions in a deterministic universe absent of metaphysically free decision-making.

## *13.1. General, Legal and Moral Responsibility*

Whereas responsibility attaches to decisions to act, it is clear that not all decisions attract the ascription of responsibility; a mere decision is necessary but not sufficient for responsibility. Consequently, the present thesis has proposed that it is only where an agent possesses the three crucial capacities of reasons responsiveness, ordinary self-control and appreciation of their actions, that they may be held responsible for their subsequent decisions. Although the focus of the present thesis has been on legal responsibility, the capacity-based theory arrived at offers in principle a *general* theory of responsibility that can be extended to apply to questions of both legal and moral responsibility. That is to say, it may be argued that the ontology of responsibility *per se* consists of a decision to act taken in possession of the three crucial capacities, this being applicable to any decision to act whether its implications are legal, moral or other.

Beginning with the general, the thesis offers an account of responsibility for human decisions and their subsequent actions, regardless of their legal or moral character. This general account provides:

> *People are responsible for the consequences of their decisions if they act whilst in possession of the capacities to respond to reason, to control their actions to accord with intentions, and to appreciate the nature and consequences of their actions and their effects in the world.*

These capacities have been more thoroughly defended in chapter nine of the present thesis, and are here rehearsed in the negative. Thus, where an individual lacks the capacity to respond to reason, they are not necessarily able to recognise and give due consideration to what may be undisputedly rational, legitimate and obvious reasons for doing or not doing any particular act. In order to weigh up the options of any decision, it is necessary to be able to both recognise what *types* of reasons are good or bad, and how respective good and bad reasons apply to weighing up the best option for action. The lack of the very capacity to do either of these things abrogates responsibility for action, because the individual cannot necessarily appreciate either the fact that their actions are prohibited / illegal / immoral, the reasons why those actions are so prohibited, or what the implications of that prohibition should be for their own behaviour.

"Ordinary" self-control has been decoupled from any explicit requirement for *conscious* control, and refers simply to the ordinary ability to accord bodily actions with what was intended in the mind – *i.e.,* I intend to grasp the cup and my hand accordingly does so. Whilst the absence of consciousness is often a sound indicator that a person lacks the capacity for ordinary self-control in that moment, this does not entail that the presence of consciousness is a necessary component of that ordinary capacity. Nonetheless, it is obvious why the absence of ordinary self-control would abrogate responsibility for subsequent actions that the individual had not intended in the event. It is for this reason, for example, that involuntary spasms and reflexes do not attract ascriptions of responsibility; people are generally not held responsible for actions occurring in any state where it is recognised that they cannot intentionally control bodily actions, including

whilst sleepwalking, during seizures and when acting as an automaton. To ascribe responsibility in such circumstances is meaningless, because the individual is unable to sufficiently control their actions to conform with proscribed behaviour, even if they wanted and intended to do so.

Finally, where an individual lacks the capacity to appreciate the nature and consequences of their actions, this means that they are not necessarily able to appreciate what will be the physical effects of their actions in the world, or their effects for other people or things, and nor, therefore, why those effects might be regarded as good or bad, legal or illegal, moral or immoral. Clearly, again, a person lacking this capacity cannot reasonably be regarded as being responsible for bringing about consequences that they could not consider and appreciate in the first place. Any prescription or proscription of behaviour becomes meaningless when a person cannot appreciate whether or not their actions will conform therewith.

The presence of the three capacities justifies ascribing responsibility for the consequences of actions in general. A brain in possession of these capacities has, *in principle*, all that it requires to form goals and intentions that are guided by good reason, to control bodily actions in order for conform with those intentions; and to appreciate the consequences of how resultant actions will bring about effects within the world. This becomes more acute when considering responsibility within the context of rules; it has been argued previously in this thesis that legal and moral rules exist fundamentally to identify desirable and undesirable conduct / outcomes, and to guide human behaviour accordingly. In this regard, H. L. A. Hart identifies that the principal purpose of the criminal law is to 'announce to society that these actions are not to be done and to secure that fewer of them are done.'[2]

Whilst this does not state that legal and moral rules *only* fulfil these ends, the purpose of guiding behaviour towards particular conduct is surely a more fundamental purpose of both legal and moral rules. When discussing actions that accord with or contravene *any* behaviour-guiding rule, the importance of the three capacities becomes ever clearer. An

[2] H. L. A. Hart, 'Prolegomenon to the principles of punishment' in Hart H. L. A. (ed.), *Punishment and Responsibility: Essays in the Philosophy of Law* (2nd ed. Oxford University Press 2008), 6.

individual must possess the capacity to be responsive to reason in order to appreciate the status of a rule and why it ought to be followed and applied. An individual similarly must be able in the practical sense to accord their actions with that rule should they so intend; and an individual must be able to appreciate the nature and consequences of their actions in order to know themselves whether those actions will follow in accordance with, or contravention of, the rule in question.

From this general account of responsibility for actions, it may readily be extended to govern questions of legal or moral responsibility. The present thesis has been predominantly concerned with the former legal responsibility, as moral responsibility is not itself a sub-requisite of legal responsibility. The law generally presumes the existence of the first two capacities in adults encapsulated within the concept of volition, and the present thesis has not raised evidence to substantively interfere with this presumption. Where the law currently requires positive proof of subjective mental states, this has been replaced with proof of the third capacity through the application of the reasonableness principle. Thus, the proposed *legal* theory of responsibility requires the prosecution to positively demonstrate that it is *reasonable to expect anybody in the same circumstances as the defendant to appreciate the nature and consequences of their actions* as they relate to the offence charged, whilst this nature is objectively defined by reference to the objective description of *mens rea* under the first limb of the hybrid test. As chapter eleven of the thesis has further demonstrated, legal defences generally operate to negate the existence or proper functioning of one or more of the three aforementioned capacities, whether the first two which are presumed to exist, or the third which must be proven to have existed to the threshold of reasonableness.

The same general capacity-based approach to responsibility can equally be applied to questions of moral responsibility. Here, an individual becomes *morally* responsible for the consequences of decisions which have a moral character if they act whilst in possession of the capacities to respond to reason, to control their actions such that they accord with intended actions, and to appreciate the nature and consequences of their actions and their effects in the world. Indeed, from starting with the general account of responsibility for actions, it becomes clear that the nature of legal and moral responsibility

is determined by the character of the rule regarding which responsibility for action is being ascribed. Put more simply, the thesis arrives at a general account of the ontology of what it is to hold a person responsible for the consequences of any action. What makes someone *legally* responsible is that the relevant actions and consequences relate to a *legal* duty or prohibition; what makes somebody *morally* responsible is that the actions and consequences in question relate to a *moral* duty or prohibition.

### *13.1.1. The Teleological Defence*

It has previously been posited in section 9.4 of this thesis, above, that the fundamental underlying function of legal and moral rules is to cause human behaviour towards identified desired ends and away from identified undesired ends. That is, rules exist to prescribe how people are expected to act, and to proscribe against how people are expected not to act; their purpose is to govern human behaviour within a social context by persuading, manipulating, coercing and / or compelling people to act or not in particular ways. This is not to say that legal and moral rules *only* consist of this function, or that there are no further conditions which identify such rules; only that these rules exist, at the most basic level, for the fundamental purpose of guiding behaviour. It follows that if such rules do not succeed in guiding behaviour then their very purpose and function has been undermined; the rules are not performing the inherent role for which they exist in the first place. Thus, a rule is undermined when a person decides to act contrary thereto; a rule that has been broken has, in that instance, failed in its fundamental object or purpose.

But not all decisions to act attract responsibility; rather, it has been argued, decisions taken where the agent possesses three crucial capacities are those for which the agent may fairly and rationally be held responsible, those capacities being: the ability to recognise and respond to good and bad reasons for acting; the ability to control bodily actions such that they conform with intended actions; and the ability to recognise the nature of one's actions as they relate to a rule (*i.e.,* their legal / moral nature) and the consequences of those actions in breaching that rule. The reason why these particular capacities are relevant is that, if it is expected that legal and moral rules are to actually cause certain behaviours, these capacities are minimally necessary and sufficient in order for it to be

reasonable to expect rules to have such a causative effect on the deciding brain. This link between the purpose of rules and subsequent punishment in guiding volitional decisions and actions enjoys a loose pedigree in Aristotle's *Nicomachean Ethics*, similarly discussed above in section 9.4 of the thesis.

As previously argued, an agent in possession of the three capacities has all that they require in principle to be able to recognise that a legal / moral rule exists in the first place, and that it is a good reason for acting in compliance therewith (*i.e.,* reasons-responsiveness); to ensure that their bodily motions do not breach that rule should they so decide (*i.e.,* ordinary self-control); and to understand that the consequences of acting one way rather than another would be to breach the said rule (*i.e.,* appreciation of the nature and consequences of actions). If the existence of a legal / moral rule is to cause a human brain to decide to act in compliance with that rule, the brain in question at least requires these capacities in order for it to be reasonably possible for that rule to influence human decisions as intended. More crucially, if the brain in question lacks any of the three capacities, then it cannot necessarily be expected to be able to identify and apply rules as good reasons for acting, to control the body to conform with such rules, and / or to appreciate how certain actions might result in breaching those rules. Thus, the capacity-based account of responsibility identifies responsible agents, that is, people who may fairly and rationally be held responsible for their actions.

This teleology of the capacity-based theory of responsibility goes deeper still, for identifying those who are responsible agents simultaneously identifies those to whom the full spectrum of punishment may rationally be applied. The underlying deterministic presumption of the thesis entails both that past (criminal) actions are caused, just as future (non-criminal) actions are similarly caused. From this perspective, punishment becomes a cause in itself of future behaviour; however, it is only *reasonable* to expect certain punishments to have the intended causative effect on *responsible* agents. For example, where the deterrent effect of punishment relies upon punishment providing a good reason for agents not to engage in criminal acts, an individual who acted in the absence of one of more of the three capacities is not necessarily able to grasp and apply the fact of punishment as a good reason for action, control their bodily motions so as comply with

that deterrence, or appreciate how certain actions might contravene that deterrence. Of course, where an individual is found not responsible for their actions, it may nonetheless be appropriate to impose certain incapacitating and / or rehabilitative sanctions in order to secure their safety and the safety of others, and the treatment of causes of their otherwise offensive actions. That notwithstanding, a further teleology of the capacity-based theory of responsibility is in identifying those agents for who the full spectrum of available punishments (in particular deterrence and expressivism) can reasonably be expected to have their intended causative effect on a future deciding brain.

Thus, the attribution of responsibility exists to identify the fact that somebody has decided to breach a given rule in circumstances where they were otherwise in possession of everything necessary and sufficient to not breach that rule, and in circumstances where (the threat of) punishment can rationally be expected to impact upon their (present and) future decisions. Given that rules exist fundamentally to affect decisions to act, this account of responsibility fulfils the teleology of rules by distinguishing between the circumstances – *i.e.,* the capacities of the deciding brain – under which it can reasonably and rationally be expected that any person would respond to the existence of a given legal / moral rule to act / refrain from acting in a particular way and, equally, respond to the additional threat of punishment for breaching that rule.

## *13.2. Free Will, Moral Responsibility, and (In)Compatibilism*

The thesis opened by adopting the fundamental *assumption* that metaphysical free will does not exist; that the classical universe and objects therein operate upon deterministic physical principles; and that this extends generally to the chemistry, biology, biochemistry, neuroscience, psychology and operations of the human brain – *i.e.,* all decisions and consequent behaviour, criminal or otherwise, are determined. The thesis therefore adopts a hard incompatibilist position vis-à-vis universal causal determinism and metaphysical free will – for former exists and, therefore, the latter cannot and does not.[3] As discussed at greater length in chapters eight and twelve of this thesis, and sections

[3] See further Robert Kane, 'The contours of contemporary free-will debates (part 2)' in Kane R. (ed.), *The Oxford Handbook of Free Will* (2nd ed. Oxford University Press 2011), 25.

8.3 and 12.1 in particular, neither the present law nor the theory of responsibility presented in this thesis require *moral* wrongdoing as a prerequisite to legal responsibility, or responsibility in general. That is to say, the capacity-based theory of responsibility which rests upon the agent possessing the three crucial capacities, *does not require* any notion of moral wrongdoing whilst remaining rational in a deterministic universe.

As is further elaborated in section 13.2, below, the capacity-based theory of responsibility is generalist, in that it provides an account of the general circumstances in which it is fair, just and rational to ascribe responsibility for a person's decisions and subsequent actions. Consequently, the capacity-based theory of responsibility can be applied within both a legal context (as in the present thesis), or a moral context. That is to say, the same capacity-based theory can be applied to ascribe both legal and moral responsibility for actions. In this regard, the present thesis is compatibilist vis-à-vis determinism and moral responsibility,[4] whilst remaining incompatibilist vis-à-vis determinism and free will.

It is submitted that the vast majority of theories that are compatibilist in this latter respect are in fact revisionist, as they typically accept the hard determinist rejection of the *metaphysical* components of free will (*i.e.,* the principle of alternative possibilities and / or the possibility for *causa sui* causes) and instead attempt to redefine "free" will according to some other criteria. Indeed, some such theories redefine free will in relation to mental capacities; however, this imprecise use of language only serves to obfuscate the real contentions at the heart of the philosophical debate, *i.e.,* the compatibility or otherwise of causal determinism and *metaphysical* free will, and the implications of denying the latter for the subsequent concept of responsibility. From here, it is further submitted that a significant body of debate has proceeded upon the assumption that the hard incompatibilist denial of free will necessitates a concurrent denial of moral responsibility – *this is not so*.

[4] See similarly, Saul Smilansky, 'Free will, fundamental dualism, and the centrality of illusion' in Kane R. (ed.), *The Oxford Handbook of Free Will* (2nd ed. Oxford University Press 2011).

The traditional hard incompatibilist denial of moral responsibility is adapted from van Inwagen:

> 'If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.'[5]

This, it is submitted, provides a solely retrospective view of the relationship between deterministic causation and responsibility – it proposes correctly that all decisions and actions are the effects of prior causes, and that no individual is responsible for the multitude of causes set in motion prior to their birth, but concludes incorrectly that no individual is therefore responsible for their present decisions and actions. The error lies in regarding a decision or action as being *solely* the effect of prior causes, as opposed to *also* being a cause in its own right; and in failing to recognise that a decision to act is something altogether qualitatively different to and greater than the sum of its causes.

With regards to the first error, the aforementioned denial of moral responsibility adopts an entirely *retrospective* view of a decision to act as being *solely* the effect of prior causes. All decisions are indeed the determined effects of a culmination of prior causes; but all decisions (to act) are also causes in themselves which proceed to have tangible effects in the world. A *prospective* view of decision-making regards the decision as a cause of further effects in the world which would not exist but for a given decision to act, howsoever that decision was caused itself; determinism is a continuous chain of cause and effect which does not stop at the caused human decision. If decisions are causes in their own right with subsequent effects, and a decision to act one way or another can have

[5] Peter van Inwagen, *An Essay on Free Will* (Clarendon Press 1983), 16. This argument is originally offered as a refutation of free will in a deterministic universe, such that to say that something is "not up to us" is to say that we have no power or control over that thing – *i.e.,* we have no power or control over the laws of the universe or prior events occurring before our birth and, therefore, no power or control over our present actions. van Inwagen's argument has been since adopted to also suggest that moral responsibility is incompatible with determinism, reading "not up to us" as meaning that we are not responsible – *i.e.,* we are not morally responsible for the laws of the universe or prior events occurring before our birth, and therefore, we are not morally responsible for our present actions.

differing effects, and those effects would not arise without the relevant decision one way or another, then decisions and actions are responsible for producing subsequent effects in the world as a matter of "but-for" causation. Whereas the retrospective argument from determinism states that one cannot be responsible for our present acts because we are not responsible for its *causes*, the prospective argument states that we can be responsible for our present acts because we are responsible for the *effects* which follow those acts and which would not otherwise have followed without those acts.

Regarding the second error, to argue that moral responsibility for action does not exist because one cannot be responsible for the prior deterministic causes of one's decisions, ignores entirely the effects in the world – the future consequences – that a decision to act causes. In this regard, a decision is plainly something different to and greater than the sum of its causes. The causes of any given decision are virtually innumerable: they include a person's genetic and epigenetic makeup, their upbringing and education, and their familial and social environment, all of which contribute to their personality and character, and executive functions such as rational thought and self-control in any given moment; they include the likes, dislikes, values, preferences, experiences and memories which inform the input of a decision-making process; they include the conversations, advertisements, memes or other subtle stimuli that may be acting as a prime on their decision-making in the moment; they include the circumstances within which a particular decision arises, and the available options for feasible responses in those particular circumstances.

At the crucial moment, all these causes of any given decision culminate in a brain holding representations of the possible decision options – the choices available to be made in response to a decision. Upon one representation reaching a threshold to become *the choice,* that decision proceeds into actions which have effects and consequences in the wider world. Thus, all those factors that deterministically culminate in causing a particular decision are transfigured into something new and qualitatively different to its causes, and which itself becomes a new cause in itself of effects in the world – an action. Without a decision, those innumerable things which cause that decision would never be transformed into an action and, thus, would never culminate in the unique effects that human action

has in the world. A decision to act is not merely the sum of its deterministic causes but results in effects that are qualitatively different from those causes, and which would not result from those causes without the decision itself occurring; b*ut for* the deciding brain, there would be no human action in the first place to which responsibility could be ascribed. To revert van Inwagen's argument:

> *If determinism is true, then our acts are the causes of subsequent consequences which, subject to the laws of nature, have effects in the future. Whilst it is not up to us what the laws of nature are, it is up to us what goes on after we are born / after we act. Therefore, the consequences of our present actions are up to us.*

The incompatibilist denial of moral responsibility presupposes that the very fact that one is not responsible for something in the past or the laws of nature transfers over to render one not responsible for actions in the present – this is the "transfer of non-responsibility" argument. However, this is only supposed and, it is submitted, the onus rests with the hard incompatibilist to actually demonstrate and exemplify this supposition. McKenna argues that counterexamples that attempt to do so run into difficulties as soon as they involve a deliberating mind within the chain of causation; that is to say, the transfer of non-responsibility argument does not appear so strong when a *decision to act* is inserted into the chain of causation. He writes, 'once attention is rightly drawn to causal sequences that begin with the onset of deliberation and end in action, it is far from clear that [the transfer of non-responsibility argument] is a defensible inference principle.'[6]

The broader point to be drawn from these errors is that the very concept of responsibility fundamentally attaches to decisions to act. Consequently, the first two propositions of the incompatibilist denial of moral responsibility are, in fact, irrelevant to the argument of whether or not responsibility can logically exist in a deterministic universe. If responsibility attaches to decisions to act, a person by definition cannot be responsible

[6] Michael McKenna, 'Saying good-bye to the Direct Argument the right way' (2008b) 117(3) *The Philosophical Review* 349, 379; see also Mark Ravizza, 'Semi-compatibilism and the transfer of non-responsibility' (1994) 71(1) *Philosophical Studies* 61.

for the laws of the universe, nor can they be responsible for any of the multitude of prior causes that lead to their decisions *and were not themselves previous decisions that person had made*. The propositions which set up the incompatibilist denial of moral responsibility are themselves irrelevant to the debate, because they argue that a person is not *actually* responsible *per se* for things (*i.e.,* the laws of the universe and the chain of prior causes) for which a person *literally cannot be* responsible in the first place, by the very definition of responsibility. Responsibility entails something more than merely causation.

### *13.2.1. Persuasion, Manipulation, Coercion and Compulsion*

The argument that responsibility attaches to decisions to act can further be elucidated by considering the effects of persuasion, manipulation, coercion and compulsion and, in particular, why the latter three are generally recognised as impacting upon an individual's responsibility for their actions whilst the former does not. Persuasion implies some manner of dialogue; more specifically, the person doing the persuading appeals to the powers of rational thought of the person being persuaded, by providing reasons and arguments to convince that person of a given thing.[7] Consequently, the fact that a person is persuaded to do a thing does not generally lessen their responsibility for so doing; indeed, Aristotle provides that rhetoric is inherently moral because it appeals to that which makes somebody peculiarly human – *i.e.,* their rationality or ability to reason – and it enables the 'true and the just [to] prevail over their opposites, in order that the decisions of our judges might be "what they ought to be".'[8] Kantian critical rhetoric acknowledges 'what it means to treat the utterances of one's self and of others as issuing forth from intrinsically valuable rational agents.'[9]

Like persuasion, manipulation does generally enter into some manner of dialogue with the manipulated individual; however, rather than engaging with another's rationality, manipulation seeks to 'circumvent or subject their rational decision-making processes,

[7] Christof Rapp, 'Aristotle on the moral psychology of persuasion' in Shields C. (ed.), *The Oxford Handbook of Aristotle* (Oxford University Press 2012), 589 – 595.
[8] Christopher Lyle Johnstone, 'An Aristotelian trilogy: Ethics, rhetoric, politics, and the search for moral truth' (1980) 13(1) *Philosophy & Rhetoric* 1, 10.
[9] Scott R. Stroud, *Kant and the Promise of Rhetoric* (Pennsylvania State University Press 2014), 200.

and… undermine or disrupt the ways of choosing that they themselves would critically endorse if they considered the matter in a way that is *lucid and free of error*.'[10] Baron identifies numerous methods of manipulation which might broadly be grouped into three categories:[11] *deception* by outright lying, misrepresentation and making false promises is quintessential manipulation, but deception may be more subtle, such as by misleading without lying, exploiting another's misunderstanding, and failing to correct their error; *pressure* such as by issuing threats the severity of which fall short of coercion, but also manipulative offers such as offering a homeless and starving person £10,000 to do *x*, and browbeating or wearing a person down until they concede from exhaustion rather than being convinced; and *exploitation* where the manipulator takes advantage of another's weaknesses or character flaws, such as eliciting another's feelings of guilt or sympathy, and exaggerating distress.

The impact of manipulation on an individual's responsibility for their actions is less clear than with persuasion, coercion and compulsion. Indeed, manipulation appears to occupy a grey area between persuasion and coercion, and the degree to which any instance of manipulation absolves a person's responsibility for subsequent actions will often be a question of proportionality between the gravity of the action concerned and the nature, force, strength, ingenuity and / or pressure of the manipulation exerted.[12] This is certainly the case with regards to pressure; whereas coercion can generally absolve responsibility for most, if not all, subsequent actions owing to the severity of the threats involved, the degree to which manipulative pressure is exculpatory will often be a matter of proportionality between the degree of pressure exerted by the particular manipulation and the gravity of the act procured thereby. Deception may often provide more complete absolution in general insofar as it induces a sufficiently exculpatory degree of misunderstanding or unawareness in the deceived. The moral consequences of exploitation are perhaps the most uncertain, and much often turns on the evaluation of the weakness or character trait that is being exploited; the exploitation of "innocent" traits

[10] Christian Coons and Michael Weber, 'Coercion, manipulation, exploitation' in Coons C. and Weber M. (eds.), *Manipulation: Theory and Practice* (Oxford University Press 2014), 35 (emphasis added).
[11] Marcia Baron, 'Manipulativeness' (2003) 77(2) *Proceedings and Addresses of the American Philosophical Association* 37, 40 – 44.
[12] See further Anne Barnhill, 'What is manipulation?' in in Coons C. and Weber M. (eds.), *Manipulation: Theory and Practice* (Oxford University Press 2014).

such as a person's lack of intelligence is typically more exculpatory to responsibility than the exploitation of a "moralised" weakness such as someone's greed or addiction.

In stark contrast to persuasion (and, to some extent, manipulation), coercion does not enter into any such dialogue; rather, the coercer seeks to secure that the coerced does a given thing through the threat of less desirable consequences should they refuse. More specifically, coercion envisages the use of threats of a particular credibility and seriousness, such that they "overpower" the will of the coerced or threaten consequences that no person is reasonably expected to withstand, such as threats of death or injury to themselves or loved ones.[13] In this regard, people do not generally succumb to threats that are simply not credible, or that threaten consequences that are less aversive than the thing that they are being threatened to do. In similar contrast to persuasion, coercion has a clear impact upon attributions of responsibility insofar as people are generally not held to be responsible for decisions to act which were coerced. The reason, again, appeals to the fact the coercion engages threats of such severity that all other options for action become effectively meaningless; the 'moral and legal justification is that requiring human beings not to yield to some threats is simply too much to ask of creatures like ourselves.'[14]

Compulsion involves the actual application of force in order to secure that a person does a given thing.[15] A person can be compelled to do a specific thing, such as when somebody grabs the arm of another and physically drags them into a location, or if somebody physically held a gun in another's hand and forced their finger on the trigger. Less obviously, a person may be compelled from doing many things generally, including certain specific things, such as in the case of a prisoner. Their incarceration physically prevents them from doing many things such as going outside, accessing the internet or meeting friends; but it also prevents (or compels) them from doing specific things like reoffending. It is also possible to speak of "internal" compulsion, for example, when a

[13] Robert Nozick, 'Coercion' in Morgenbesser W. (ed.), *Philosophy, Science and Method: Essays in Honor of Ernest Nagel* (St. Martin's Press 1969); Harry G. Frankfurt, 'Coercion and moral responsibility' in Frankfurt H. G. (ed.), *The Importance of What We Care About: Philosophical Essays* (Cambridge University Press 1998); Mitchell N. Berman, 'The normative functions of coercion claims' (2002) 8(1) *Legal Theory* 45.
[14] Stephen J. Morse, 'Moral and legal responsibility and the new neuroscience' in Illes J. (ed.), *Neuroethics: Defining the Issues in Theory, Practice, and Policy* (Oxford University Press 2006), 38.
[15] Denis G. Arnold, 'Coercion and moral responsibility' (2001) 38(1) *American Philosophical Quarterly* 53.

person moves due to an uncontrollable reflex or tremor, through epileptic fit or a hypoglycaemic state.[16] Like coercion, people are not held responsible for actions that are compelled, often appealing to the way in which compulsion bypasses the individual's ability to choose altogether; their choice becomes entirely irrelevant, because their actions are procured by force irrespective of their will.

Coercion has received the greatest philosophical interest, perhaps because persuasion does not necessarily impact upon questions of responsibility, whilst a person's being compelled to act is often all the rarer and more obvious. Two points may be drawn from the literature on coercion which, it is argued, can extend equally to persuasion and compulsion. First, it is submitted that coercion is concerned with threats which cause a person to act otherwise than they would. In this regard, Nozick's discussion of coercion emphasises the effect of coercion on changing a person's choice;[17] meanwhile, Frankfurt writes that 'in submitting to a threat, a person invariably does something which he does not really want to do.'[18] To say that a person's decision to do *x* was coerced, therefore, is to say that they would not otherwise have decided to do *x* but for the coercive threat.

The same argument may be extended to persuasion, manipulation and compulsion. Again, if a person decides to do *x* and another comes along and offers a number of persuasive arguments to do *x*, it would not be correct to describe the first individual as having been persuaded to do *x*. They may have been encouraged or further motivated by the persuasive arguments, but the decision to do *x* (for example, instead of *y*) was not *persuaded*; to be persuaded is to be moved by argument from one position to another. Similarly, a person cannot be said to have been manipulated into doing *x* when they had already decided to do so. And, again, the same with regards to compulsion, albeit somewhat less obviously:

[16] Stephen J. Morse, 'Causation, compulsion, and involuntariness' (1994) 22(2) *Bulletin of the American Academy of Psychiatry and the Law* 159.
[17] Nozick (1969).
[18] Frankfurt (1998), 43; see also Michael J. Murray and David F. Dudrick, 'Are coerced acts free?' (1995) 32(2) *American Philosophical Quarterly* 109.

> *Sally goes into a room to read her book and, unbeknownst to her, James locks the door behind her for an hour and then unlocks it again. After one-and-a-half hours of reading, Sally leaves the room.*

It does not seem natural to suggest that Sally was compelled to stay in the room even during the time when the door was locked; this fact had no bearing on what she wanted or chose to do. If, however, Sally had attempted to leave after thirty minutes to find that she could not, she would from that point be compelled to stay in the room. What has changed in the latter scenario is that, but for the compulsion, Sally would not stay in the room whereas, in the first scenario, she would stay in the room regardless. Thus, to say that a person's decision to *x* was persuaded, manipulated, coerced or compelled is to say that that person would have decided to not *x*, but for the persuasion / manipulation / coercion / compulsion.

Nozick further famously introduces a success condition into his consideration of coercion, such that if the coerced individual does not act as they have been threatened, then there has been no coercion of their decision to act.[19] Where a person decides to do *x* and another threatens them unless they do *y*, if that first person proceeds to do *x* regardless of the threat, then the threat has failed; this may be a case of attempted coercion, but no actual coercion of a decision to act has occurred. This argument may, again, be extended to persuasion, manipulation and compulsion; where a person decides to do *x* and another persuades / manipulates / compels them to do *y*, if that first person continues to do *x* then the persuasion / manipulation / compulsion has not occurred as a matter of fact. From this discussion, it becomes clear that the differing effects on ascriptions of responsibility for actions that are persuaded, manipulated, coerced or compelled has nothing to do with the *causal* role that persuasion / manipulation / coercion / compulsion plays in a person's decision-making. For any decision to *x* to have been persuaded… *etc.,* the acts of persuasion… *etc.,* must have been both necessary and sufficient in the circumstances of the particular decision to secure *x*.

[19] Nozick (1969), 441 – 445.

It is submitted that what distinguishes persuasion… *etc.*, therefore, is the *manner* in which they operate on a person's decision to act and, more specifically, how they impact upon the three capacities of reasons responsiveness, ordinary self-control and appreciation of the nature and consequences of actions. As argued above, persuasion explicitly engages with another's rationality in providing reasons and arguments for a person to consider and apply to their decision. Thus, persuasion does not operate by overwhelming or bypassing a person's capacities at all but, quite the opposite, by engaging those capacities. A person who is merely persuaded to do *x* is still responsible for so doing, not least where *x* breaches some legal or moral rule which should have provided good reason not to *x* despite the persuasion. They remain responsible because, notwithstanding the persuasion, they decided to do *x* in circumstances where they had the capacity to recognise that *x* was a prohibited legal / moral wrong and that this was a good reason not to *x*, where they had the capacity to control their bodily actions to not do *x* had they so decided, and where they had the capacity to appreciate the nature and consequences that doing *x* would have in breaching the legal / moral prohibition.

The way in which manipulation operates upon a person's decision-making capacities depends on the type of manipulation concerned. Manipulation through deception entails some form of *covert* influence; the manipulated individual 'either has no knowledge of, or does not understand, the ways in which [the manipulation] affects his choices.'[20] Then again, it does not seem that responsibility is vitiated in every instance that a person acts because of a lie, or every case when they do not appreciate the underlying intentions of another. Rather, it is submitted, in order to excuse responsibility for acting, any deception must have so affected the individual's capacity to appreciate the nature and consequences of their actions. For example, if *A* wants *B's* bag for his personal gain and deceives *C* into stealing it for him, it makes little difference to *C's* responsibility if he is convinced with the lie; "*B* is a bad person who steals from the poor and does not pay his taxes."

However, if *A* lies and tells *C* that the bag is in fact his which *B* took from him earlier, *C* might more readily be excused for stealing the bag. In particular, this lie, if credible and

[20] Alan Ware, 'The conception of manipulation: Its relation to democracy and power' (1981) 11(2) *British Journal of Political Science* 163, 165.

believed, has specifically targeted *C*'s ability to appreciate that his actions are stealing and therefore wrongful; for all he believes (due to the lie), *C* is doing a good deed of recovering *A'*s stolen property. Mills writes that deception attempts to 'change another's beliefs and desires by offering her bad reasons, *disguised as good*, or faulty arguments, *disguised as sound*.'[21] When successful, these reasons and arguments lead the deceived individual to believe that the nature or consequences of their actions is one thing when, in fact, it is another.

Manipulation by pressure is less obvious; whereas the impact of coercion on responsibility is obvious because it consists of threats of such gravity that no reasonable person is reasonably expected to resist, pressure may consist of all such threats which do not otherwise amount to coercion, as well as pressurising offers, and other forms such as browbeating and wearing another down. Because the impact of these actions is not obvious, they will always require a closer inspection in practice. Generally speaking, whether or not a person is responsible for a decision to act made under pressure will depend on both the nature and severity of the pressure exerted (and whether reasonable people are generally expected to resist or submit to that pressure), and the gravity of the thing that the individual is being pressured to do (considered in proportion to the pressure itself).

The question with pressure by threats will be whether, in the circumstances of a particular decision, the threats were of sufficient severity to overwhelm the individual's responsiveness to good reason not to perform the threatened action. The question with pressure exerted over time, such as browbeating, will be whether the pressure was of sufficient severity to overwhelm the individual's ordinary self-control which previously prevented them from acquiescing. The law recognises partial defences of diminished responsibility and loss of control in such circumstances, discussed in section 11.3.6 of this thesis above; and it is submitted that morality equally may recognise the mitigating effects of manipulative pressure on responsibility. As highlighted, however, this will often be mitigatory as opposed to entirely exculpatory (contrasting with coercion and

[21] Claudia Mills, 'Politics and manipulation' (1995) 21(1) *Social Theory and Practice* 97, 100 (emphasis added).

compulsion), due to the inherent question of proportionality between the pressure exerted and the seriousness of the act being induced.

Manipulation by exploiting character traits or weaknesses is, again, less obvious.[22] If *A* is prone to stealing or violence, and *B* exploits this by giving them reasons and arguments to steal or commit violence, the intuition generally follows that *A* is not absolved of responsibility for their actions. Conversely, *A* may more readily be absolved if they are particularly vulnerable, young, or unintelligent, and *B* exploits this to securing their action. This might suggest that the question turns on some moral judgment given to the characteristic or weakness that is being exploited – whether it is an "innocent" or "moralised" trait. However, this is not necessarily so; for example, *A* could be particularly charitable or sympathetic and it is these qualities which are exploited – perhaps *B* appeals to *A*'s sympathy with a suitable story and then encourages them to steal in order to donate to good causes, exploiting their charitable nature. Whilst charitableness and sympathy are unquestionably good characteristics, their exploitation does not necessary provide absolution to *A*'s subsequent act of theft.

Rather, it is submitted that the kind of exploitation relevant to responsibility involves that which undermines the individual's capacity to appreciate the nature and consequences of their actions. This is why the exploitation of the vulnerable or unintelligent may be treated differently; when *B* manipulates *A* by exploiting their misunderstanding or misapprehension of a situation, this kind of exploitation, or deception by omission (*i.e.,* permitting *A* to act under their false pretences), causes *A* to act without being able to appreciate the true nature and consequences of their actions.

Coercion operates through the use of threats of such credibility and gravity that they overpower the will of the coerced, leaving no reasonable option but to act in concordance with the threat that was made. In this respect, it is submitted that coercion operates

[22] For example, see Joel Rudinow, 'Manipulation' (1978) 88(4) *Ethics* 338, 346; Ruth R. Faden and Tom L. Beauchamp, *A History and Theory of Informed Consent* (Oxford University Press 1986), 366.

specifically on the capacity to recognise and respond to good or bad reasons for acting. Gert proposes that coercion offers an "unreasonable incentive", by which is meaned:

> 'An incentive is unreasonable if it would be unreasonable to expect any rational man in that situation not to act on it.[23] A reasonable incentive is an incentive that is not unreasonable. It would be unreasonable to expect any rational man not to act on certain consequences only if those consequences always provide motives for all rational men. If consequences are such that they provide motives only generally, but not always, to all rational men, then they cannot be unreasonable incentives for it would not be unreasonable to expect some rational man not to act because of them. Consequences which involve the gaining of a good only generally, but not always, provide motives to rational men. Therefore, consequences which only involve the gaining of a good cannot be unreasonable incentives. Consequences which involve the avoiding of an evil always provide motives to all rational men, for all rational men must seek to avoid any evil – unless they have a reason. This means that the belief that they will avoid an evil always serves as a motive for all rational men. Of course, not all consequences that involve the avoiding of an evil will be unreasonable incentives. The evils must be significant; usually only death, severe and prolonged pain, serious disability, and extensive loss of freedom will be unreasonable incentives. Only serious evils such as these provide motives that make it unreasonable to expect any rational man not to act on.'[24]

Unreasonable incentives override the capacity to be responsive to good reason because the severity of the threatened consequence is so great that no other reason can ever be sufficiently good to refrain from acquiescing to the threat. If a man is threatened at gunpoint unless he steals a woman's handbag, it would make no difference to his

[23] Bernard Gert, 'Coercion and freedom' in Pennock J. R. and Chapman J. W. (eds.), *Coercion* (Taylor & Francis 1973), 34; Gert notes that this definition of an unreasonable incentive is close to what Stanley I. Benn calls a reasonable intention.
[24] Gert (1973), 34 – 35.

subsequent decision if another came along and provided every good reason why he should not commit the act. Against the threat of death, objections such as *theft is wrong, the woman is vulnerable, it is against the law, he may go to prison, etc.* have no impact upon decision-making. Although the man may still continue to recognise that these would be good reasons not to steal, the existence, credibility and severity of the threat – the unreasonable incentive – effectively prevents him from being able to apply these good reasons in deciding how to act; only the threat matters.

Compulsion might be conceived in one of at least two ways. It is submitted, on the one hand, that compulsion is best understood as interfering with an individual's capacity for ordinary self-control. A person who decides to act in one way but is compelled to act in another does not possess sufficient self-control to conform their bodily motions with the necessary actions to meet their decision as a matter of practical reality. Whether somebody is physically dragged into one location such that they cannot resist even if they wished; whether a person's incarceration prevents them practically from being able to carry out their wish to attack a rival; and where a person's epilepsy, hypoglycaemic state or automatism causes them to carry out a particular act; the essence of compulsion is that the person lacks the self-control to bring about the acts required to conform with their wishes. On the other hand, as indicated previously in this section, compulsion may be understood as bypassing a person's choice altogether; they are compelled to act regardless of whatever choice they might make.

*

The present discussion has aimed to further demonstrate that the salient "thing" which fundamentally attracts ascriptions of responsibility is the decision to act; that is, the decision to do a particular thing that immediately precedes and thus causes the bodily motions to bring that decision into physical action. What is more, the fact that these decisions are themselves caused is immaterial; indeed, the underlying assumption of universal determinism necessarily requires that these decisions, like anything else, result from a causal process. Through the comparison of persuasion, manipulation, coercion and compulsion, it is possible to appreciate that, when successful, each of these processes has

the same *causal* effect on a person's decision to act; *i.e.,* instances of successful persuasion / manipulation / coercion / compulsion provide both necessary and sufficient causes of a person's decision.

Despite successful persuasion… *etc.* exerting an identical *causal* relationship on person's choices, these processes readily have different implications for responsibility. At least, persuasion is rarely (if ever) exculpatory; manipulation might be exculpatory depending upon the circumstances; and coercion and compulsion are generally always exculpatory. Thus, through a comparison of the different operation of these processes on a person's decision and resultant implications for their responsibility, it is argued that it is the *manner* in which these processes engage with an individual's capacities for reasons responsiveness, ordinary self-control and appreciating the nature of their actions that is relevant to responsibility. As these are the capacities required in order for a person to be responsible for their choice, this further highlights the central significance of the decision to act to the concept of responsibility generally.

### *13.2.2. Frankfurt and Decisions*

Further support for the argument that responsibility necessarily attaches to decisions to act may be drawn from the famous counterexamples by Harry Frankfurt, originally developed to refute the claim that moral responsibility requires free will in the sense of it being metaphysically possible for an agent to act in alternative ways within a given set of circumstances / causes (*i.e.,* the principle of alternative possibilities).[25] The current discussion does not propose to interfere with Frankfurt's central argument – indeed, as the present thesis builds upon an incompatibilist assumption of determinism without the possibility of free will, it is neither warranted nor necessary to seek to rescue the principle of alternative possibilities from Frankfurt's counterexamples. Rather, it is submitted that these counterexamples further demonstrate the central importance of *choice* as the crucial

[25] Harry G. Frankfurt, 'Alternate possibilities and moral responsibility' (1969) 66(23) *Journal of Philosophy* 829.

factor to which the concept of responsibility itself attaches, notwithstanding that any such choice results from a deterministic chain of causation.

To begin, Frankfurt introduces Jones who 'decides for reasons of his own to do something,' and is then threatened with a penalty so harsh that any reasonable person would submit to the threat, unless he does precisely the thing he has already decided to do. Frankfurt then offers a number of scenarios, asking in each case whether or not Jones is morally responsible for his actions.

> 'One possibility is that $Jones_1$ is not a reasonable man: he is, rather, a man who does what he has once decided to do no matter what happens next and no matter what the cost. In that case, the threat actually exerted no effective force upon him. He acted without any regard to it, very much as if he were not aware that it had been made. If this is indeed the way it was, the situation did not involve coercion at all. The threat did not lead $Jones_1$ to do what he did. Nor was it in fact sufficient to have prevented him from doing otherwise: if his earlier decision had been to do something else, the threat would not have deterred him in the slightest. It seems evident that in these circumstances the fact that $Jones_1$ was threatened in no way reduces the moral responsibility he would otherwise bear for his act.'[26]

In this first counterexample, $Jones_1$' decision to do *x* was entirely unaffected by the threat; whether he happens to choose to do the thing that is threatened or something different, the threat has no coercive effect on that decision. In particular, $Jones_1$' capacities to recognise and apply good and bad reasons in his decision-making and to control his bodily motions to conform with his intentions each remained intact and operable when he made his decision; $Jones_1$ is therefore responsible for his decision to do *x*.

> 'Another possibility is that $Jones_2$ was stampeded by the threat. Given that threat, he would have performed that action regardless of what decision he had already made. The threat upset him so profoundly, moreover, that he

[26] *Ibid.*, 831.

> completely forgot his own earlier decision and did what was demanded of him entirely because he was terrified of the penalty with which he was threatened. In this case, it is not relevant to his having performed the action that he had already decided on his own to perform it. When the chips were down he thought of nothing but the threat, and fear alone led him to act. The fact that at an earlier time $Jones_2$ had decided for his own reasons to act in just that way may be relevant to an evaluation of his character; he may bear full responsibility for having made *that* decision. But he can hardly be said to be morally responsible for his action. For he performed the action simply as a result of the coercion to which he was subjected. His earlier decision played no role in bringing it about that he did what he did, and it would therefore be gratuitous to assign it a role in the moral evaluation of his action.'[27]

Contrary to Frankfurt, it is submitted that $Jones_2$ is responsible for his decision to do *x*. Notwithstanding Frankfurt's description of the threat – *i.e.,* that it stampeded $Jones_2$ who completely forgot his own previous decision to do *x*, and caused him to do *x* out of fear of the threat – the threat itself has no *necessary* effect on $Jones_2$ actually doing *x*. If the threat had not occurred, $Jones_2$ would still have proceeded to do *x* based on his own decision; if, after the threat, $Jones_2$ had also recalled his own decision to do *x*, he would still have proceeded to do *x*; if the threat had subsequently been removed, $Jones_2$ would still have proceeded to do *x*; if, after doing *x*, Jones then recalled his decision prior to the threat, he will be perfectly satisfied at having done *x* in the event. Frankfurt does concede that it would be fair to hold Jones responsible for his decision prior to the threat, but not for his actions after. It is submitted that this concession lends some further support to the argument that responsibility attaches to a *decision* to act, *i.e.,* the decision for which Frankfurt himself permits us to hold Jones responsible. However, his error lies in the fact that, where responsibility-absolving coercion involves a credible and significant *threat* which is both *necessary* and *sufficient* to causing a person to do *x*, the threat in this counterexample is only sufficient but not necessary to cause $Jones_2$' actions. The threat

[27] *Ibid.,* 832.

is not, by definition, coercive, because it never changes $Jones_2$' decision to act from what it otherwise would have been.

> *Suppose $Jones_{2a}$ is a gangster and Black is his boss. They decide together to go and take out some rivals who are stealing business and threatening $Jones_{2a}$ and Black's family; both $Jones_{2a}$ and Black each therefore want to proceed with taking out their rivals, both for their own safety and to make more money by taking the rivals' business. Neither $Jones_{2a}$ nor Black is coercing the other to embark upon this course of action, which will be mutually beneficial to both of them. They get in their car, drive to their rivals' hideout, roll down the windows, and the shooting begins. $Jones_{2a}$ draws his gun but then freezes, momentarily stunned by the sudden loud cracks of the guns; it is not that he has changed his mind about taking out his rivals, but that he has momentarily forgotten himself and what he is doing in the furore of all the shooting. Black then points his gun at $Jones_{2a}$ and says, "if you don't start shooting, I'll shoot you." $Jones_{2a}$ snaps out of his daze and, knowing Black to be deadly serious when he makes such threats and very conscious that Black will be true to his word, $Jones_{2a}$ starts shooting and kills one of his rivals.*

Here, it is submitted that $Jones_{2a}$ is responsible for shooting and killing his rival; we would not say that $Jones_{2a}$ is absolved of responsibility in the circumstances because of the threat from Black. If $Jones_{2a}$ had not been momentarily stunned at all, he would have simply begun shooting and killed his rival. If, instead of the threat, $Jones_{2a}$ had been snapped out of his daze by a bullet hitting perilously close to his head, his response would be to begin shooting in concurrence with his intentions as they were before he was momentarily frozen. Equally, if Black has simply waited for $Jones_{2a}$ to come around from his daze naturally, upon re-realising his situation and intentions therein, $Jones_{2a}$ would still have started shooting at his rivals. Crucially, Black's threat never coerced $Jones_{2a}$ to act differently to what he had actually already decided and, therefore, was not coercion.

> *Suppose Jones$_{2b}$ is a gangster and Black is his boss, and they decide together to go and take out their rivals following the same scenario as Jones$_{2a.}$ This time, however, when Jones$_{2b}$ draws his gun and freezes, it is because he realises the horror of the situation and changes his mind. He sees one of Black's bullets hit a rival and is suddenly confronted with the reality of taking another life. Jones$_{2b}$ puts down his gun and tells Black he wants to play no part in a massacre. Black then points his gun at Jones$_{2a}$ and says, "if you don't start shooting, I'll shoot you." Jones$_{2b}$ still does not want to take part but, knowing Black to be deadly serious when he makes such threats and very conscious that he will be true to his word, Jones$_{2b}$ picks his gun back up, starts shooting and kills one of his rivals.*

It is more readily clear, it is submitted, that Jones$_{2b}$ bears less responsibility for his actions than Jones$_{2a.}$ Black's threat is sufficient to cause both Jones$_{2a}$ and Jones$_{2b}$ to start shooting and kill one of their rivals; but the threat is only *necessary* for Jones$_{2b.}$ If Black does not issue the threat, Jones$_{2a}$ will return from his momentary daze and begin shooting anyway, whereas Jones$_{2b}$ will not shoot to kill one of his rivals because he has witnessed the horror of taking a life and does not want to participate in a massacre; he has put his gun down before any threat is issued. Jones$_{2b}$ has decided not to shoot when Black issues his threat, and would not have proceeded to pick up his gun and shoot but for that threat. The threat is therefore coercive, and Jones$_{2b}$ appreciably bears less responsibility in the circumstances.[28]

Jones$_{2a}$' scenario is that which parallels with Jones$_2$ whom, Frankfurt argued, we do not hold responsible for their actions. Conversely, it is submitted that Jones$_{2a}$ (and, therefore Jones$_2$) clearly bears greater responsibility for their actions than Jones$_{2b}$. On any account, Jones$_2$ and Jones$_{2a}$ intended, desired, and would have brought about the actions which followed the threat in any event, because those actions concurred with something they had decided to do with their capacities fully intact. Knowing what Jones$_2$ and Jones$_{2a}$ had

[28] That is not to say that Jones$_{2b}$ is not still responsible for his behaviour of being a gangster in the first place, nor for going to his rivals armed with a gun and the intention to seek out confrontation. He may remain responsible for his contribution to bringing about the circumstances of the attack without specifically being responsible for having fired his gun in this particular instance, in contrast to Jones$_{2a.}$

decided to do regardless of the threat, it seems perfectly appropriate that they be subjected to punishment. When $Jones_{2b}$'s capacities were fully intact, however, he decided to do something opposite to what he actually proceeded to do in the event; when $Jones_{2b}$ decided to proceed to shoot, his capacities were not fully intact as a consequence of the coercive threat, which caused him to change his mind and do something that he would not otherwise have done. Thus, $Jones_2$ and $Jones_{2a}$'s decisions to act were arrived at with capacity, whilst $Jones_{2b}$'s was not.

> 'Now consider a third possibility. $Jones_3$ was neither stampeded by the threat nor indifferent to it. The threat impressed him, as it would impress any reasonable man, and he would have submitted to it wholeheartedly if he had not already made a decision that coincided with the one demanded of him. In fact, however, he performed the action in question on the basis of the decision he had made before the threat was issued. When he acted, he was not actually motivated by the threat but solely by the considerations that had originally commended the action to him. It was not the threat that led him to act, though it would have done so if he had not already provided himself with a sufficient motive for performing the action in question.'[29]

Frankfurt permits that $Jones_3$ is responsible for his actions where $Jones_2$ was not; however, it is clear that the same arguments can be applied to both. Crucially, neither $Jones_2$ nor $Jones_3$ were caused to *change their minds* as a result of the threat; the threat was not necessary to precuring their subsequent actions and was not, therefore, coercive. The fact that $Jones_2$ was stampeded by the threat whereas $Jones_3$ was not is immaterial, because both were uncoerced when they *decided to act* prior to the threat, whilst the threat had no necessary causal effect on the actions which subsequently followed that decision to act.

> 'Suppose someone – Black, let us say – wants $Jones_4$ to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until $Jones_4$ is about to make up his mind what to do, and he does nothing unless it is

[29] Frankfurt (1969), 832.

> clear to him (Black is an excellent judge of such things) that Jones$_4$ is going to decide to do something *other* than what he wants him to do. If it does become clear that Jones$_4$ is going to decide to do something else, Black takes effective steps to ensure that Jones$_4$ decides to do, and that he does do, what he wants him to do. Whatever Jones$_4$'s initial preferences and inclinations, then, Black will have his way… Now suppose that Black never has to show his hand because Jones$_4$, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. In that case, it seems clear, Jones$_4$ will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to take steps to ensure that he do it. It would be quite unreasonable to excuse Jones$_4$ for his action, or to withhold the praise to which it would normally entitle him, on the basis of the fact that he could not have done otherwise. This fact played no role at all in leading him to act as he did. He would have acted the same even if it had not been a fact. Indeed, everything happened just as it would have happened without Black's presence in the situation and without his readiness to intrude into it.'[30]

Frankfurt's purpose with this fourth counterexample is to demonstrate resolutely that moral responsibility may follow even in circumstances where the agent could not have acted otherwise than they did which, as previously stated, is not contended in this present discussion. Frankfurt suggests that the steps taken by Black to secure Jones$_4$'s compliance are immaterial to the point;[31] however, it is submitted, this counterexample may be adapted to demonstrate how it is precisely the *manner* of Black's intervention and its impact upon Jones$_4$'s decision to act that is relevant to its implications for responsibility. Jones$_4$ is clearly responsible because his decision to act was taken with his capacities present, whilst Black did not intervene with this decision to act at all.

[30] *Ibid*., 835 – 836.
[31] *Ibid*., 835.

*Suppose that Jones$_{4a}$ has no desire or intention to kill Black's rival, Brown, and has in fact decided not to do so. Black kidnaps Jones$_{4a}$'s daughter and threatens to kill her unless Jones$_{4a}$ kills Brown. The threat is credible and clearly of grave severity, such that Jones$_{4a}$ can see no other option but to comply. Jones$_{4a}$'s colleague, White, tries to warn him that there is a serious chance that he will be caught and sent to prison, or even killed during the attempt; but no reason could distract Jones$_{4a}$ from the immediate, credible and grave peril that his daughter is in, and what he is able to do in the circumstances to guarantee her safety. Consequently, Jones$_{4a}$ decides to kill Brown and proceeds to do so.*

Jones$_{4a}$ acts under coercion, which has the same impact on his responsibility as would compulsion also; that is, Jones$_{4a}$ is not morally[32] responsible for killing Brown. Black has made what is clearly a credible and significantly serious threat which overwhelms Jones$_{4a}$'s capacity to consider, weigh and apply any good reason to do other than to kill Brown. What is more, Black's threat was both necessary and sufficient for securing that Jones$_{4a}$ decides to kill Brown.

*Suppose that Jones$_{4b}$ has no desire or intention to kill Black's rival, Brown, and has in fact decided not to do so. But Black is a masterful persuader; he does not use force or issue threats of violence, but is always able to present the perfect set of reasons for doing a thing, and can match any objections with persuasive counterexamples. Jones$_{4b}$ resists the case put by Black but, true to form, Black can always produce the perfect counterargument to Jones$_{4b}$'s resistance. After much discussion, Jones$_{4b}$ is convinced to kill Brown and proceeds to do so.*

Whilst we may readily hold Black morally responsible for his role in persuading Jones$_{4b}$ to kill Brown, Jones$_{4b}$ is not readily absolved of responsibility. What is more, we can see

[32] Assuming the condition that the threat to Jones$_{4a}$'s daughter was so immediate that he could not reasonably have sought the aid of the police first before being presented with the opportunity to kill Brown.

that this is not due to any *causal* difference between coercion and persuasion; for both $Jones_{4a}$ and $Jones_{4b,}$ the respective coercion and persuasion was both necessary and sufficient to cause them to decide to kill Brown. Thus, it is in the different manner in which coercion and persuasion operate on a decision to act that is relevant to the ascription of responsibility; the capacity for $Jones_{4a}$ to weigh and respond appropriately to good reasons not to kill Brown was overwhelmed by the coercive threat, whereas the capacities of $Jones_{4b}$ were intact when he decided to act – indeed, Black specifically *interacted* or *engaged* with, rather than overpowered, those capacities in order change $Jones_{4b}$'s decision.

### *13.2.3. Pereboom, Responsibility and Determinism*

It is further possible still to highlight the crucial importance of a decision to act by engaging with Pereboom's famous four-case manipulation argument against moral responsibility in a deterministic universe.[33] The manipulation argument presents four cases, the first of which contains such radical features of manipulation that the reader intuits that the agent in that case is not responsible for their actions; the cases proceed to the fourth which describes the agent in terms of normal causal determinism, challenging the reader to identify a 'relevant and principled difference between any two adjacent cases' which justify why the agent might be responsible in one case but not the other. Crucially, Pereboom focuses each case on an agent who satisfies several of the most important conditions that are argued by various compatibilists as being necessary for moral responsibility.

The agent in each case is Plum who decides to kill White in order to obtain some personal advantage, and is successful in so killing them; the decision whether or not to kill White is under examination. At all material times, Plum meets various compatibilist conditions.[34] Thus, Plum meets Humean requirements that his decision both fits with his character and is not the result of some *irresistible* urge; 'since for Plum it is generally true that selfish reasons weigh heavily – too heavily when considered from a moral point of

[33] Derk Pereboom, *Free Will, Agency, and Meaning in Life* (Oxford University Press 2014), Ch. 4.
[34] *Ibid.,* 75.

view – while in addition the desire that motivates him to act is nevertheless not irresistible for him.'[35] Plum meets conditions proposed by Frankfurt for his first order desire (*i.e.,* his will to murder White) to conform with his second order desire regarding which effective desires he has; 'he wills to murder her, and he wants to will to do so.'[36] Plum is described as being responsive to reasons as advocated by Fischer and Ravizza, and his desires can be appropriately modified by good or bad reasons to act; notwithstanding that he wants to kill White, Plum would not decide to kill her 'if he believed that the bad consequences for himself… would be more severe than he actually expects them to be.'[37] Plum meets Wallace's related condition of being able to understand and apply moral reasons to regulate his actions; although he is generally quite egoistic and motivated by selfish reasons, 'when egoistic reasons that count against acting morally are weak, he will typically act for moral reasons instead.' [38] Plum further possesses the capacity to 'reflexively… revise and develop his moral character and commitment over time, and for his actions to be governed by those moral commitments' as advocated by Mele[39] and Haji.[40]

> 'Case 1: A team of neuroscientists has the ability to manipulate Plum's neural states at any time by radio-like technology. In this particular case, they do so by pressing a button just before he begins to reason about his situation, which they know will produce in him a neural state that realizes a strongly egoistic reasoning process, which the neuroscientists know will deterministically result in his decision to kill White. Plum would not have killed White had the neuroscientists not intervened, since his reasoning would then not have been sufficiently egoistic to produce this decision. But at the same time, Plum's effective first-order desire to kill White conforms

[35] *Ibid.,* citing James A. Harris, *Of Liberty and Necessity: The Free Will Debate in Eighteenth-Century British Philosophy* (Oxford University Press 2005), Ch. 3.
[36] *Ibid.,* citing Harry G. Frankfurt, 'Freedom of the will and the concept of a person' (1971) 68(1) *Journal of Philosophy* 5.
[37] *Ibid.,* citing John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge University Press 1998).
[38] *Ibid.,* citing R. Jay Wallace, *Responsibility and the Moral Sentiments* (Harvard University Press 1994).
[39] *Ibid.,* citing Alfred R. Mele, *Autonomous Agents: From Self-Control to Autonomy* (Oxford University Press 1995).
[40] *Ibid.,* citing Ishtiyaque Haji, *Moral Appraisability: Puzzles, Proposals, and Perplexities* (Oxford University Press 1998).

> to his second-order desires. In addition, his process of deliberation from which the decision results is reasons-responsive; in particular, this type of process would have resulted in Plum's refraining from deciding to kill White in certain situations in which his reasons were different. His reasoning is consistent with his character because it is frequently egoistic and sometimes strongly so. Still, it is not in general exclusively egoistic, because he sometimes successfully regulates his behavior by moral reasons, especially when the egoistic reasons are relatively weak. Plum is also not constrained to act as he does, for he does not act because of an irresistible desire – the neuroscientists do not induce a desire of this sort.'[41]

Pereboom argues intuitively that Plum is not responsible for killing White in Case 1; he offers one potential explanation as resting in the fact that Plum's decision to kill was ultimately causally determined by the neuroscientists' intervention, a matter entirely beyond his control and but for which he would not have decided to kill White in this particular instance.

> 'Case 2: Plum is just like an ordinary human being, except that a team of neuroscientists programmed him at the beginning of his life so that his reasoning is often but not always egoistic (as in Case 1), and at time strongly so, with the intended consequence that in his current circumstances he is causally determined to engage in the egoistic reasons-responsive process of deliberation and to have the set of first and second-order desires that result in his decision to kill White. Plum has the general ability to regulate his actions by moral reasons, but in his circumstances, due to the strongly egoistic nature of his deliberative reasoning, he is causally determined to make his decision to kill. Yet he does not decide as he does because of an irresistible desire. The neural realization of his

[41] Pereboom (2014), 76 – 77.

> reasoning process and of his decision is exactly the same as it is in Case 1 (although their causal histories are different).'[42]

Here, again, Pereboom argues that the intuition is against holding Plum morally responsible for killing White; and from these intuitions on Cases 1 and 2, Pereboom submits that the compatibilist conditions for moral responsibility are insufficient. He adds further, and crucially, that it 'would seem unprincipled to claim that [in Case 2], by contrast with Case 1, Plum is morally responsible *because the length of time between the programming and his decision is now great enough*. *Whether the programming occurs a few seconds before or forty years prior to the action seems irrelevant to the question of his moral responsibility.*'[43]

> 'Case 3: Plum is an ordinary human being, except that the training practices of his community causally determined the nature of his deliberative reasoning processes so that they are frequently but not exclusively rationally egoistic (the resulting nature of his deliberative reasoning processes are exactly as they are in Cases 1 and 2). This training was completed before he developed the ability to prevent or alter these practices. Due to the aspect of his character produced by this training, in his present circumstances he is causally determined to engage in the strongly egoistic reasons-responsive process of deliberation and to have the first and second-order desires that issue in his decision to kill White. While Plum does have the general ability to regulate his behavior by moral reasons, in virtue of this aspect of his character and his circumstances he is causally determined to make his immoral decision, although he does not decide as he does due to an irresistible desire. The neural realization of his deliberative reasoning process and of the decision is just as it is in Cases 1 and 2.'[44]

[42] *Ibid.,* 77.
[43] *Ibid.,* 78 (emphasis added).
[44] *Ibid.*

Case 3 might be thought of as reflecting the case of a "rotten social background" where, for example, a person is raised in a violent / abusive / impoverished background where crime and other immoral activities are normalised and encouraged. Pereboom argues that, in order to argue that Plum is responsible in Case 3, it is necessary to identify some salient feature of the circumstances which explain that responsibility in Case 3 but not in Case 2. He argues that no such feature can be identified and that, as causal determinism was sufficient to absolve Plum of responsibility in Cases 1 and 2, so it follows in case 3 on the same grounds. Finally:

> 'Case 4: Everything that happens in our universe is causally determined by virtue of its past states together with the laws of nature. Plum is an ordinary human being, raised in normal circumstances, and again his reasoning processes are frequently but not exclusively egoistic, and sometimes strongly so (as in Cases 1 – 3). His decision to kill White issues from his strongly egoistic but reasons-responsive process of deliberation, and he has the specified first and second-order desires. The neural realization of Plum's reasoning process and decision is exactly as it is in Cases 1 – 3; he has the general ability to grasp, apply, and regulate his actions by moral reasons, and it is not because of an irresistible desire that he decides to kill.'[45]

If the intuition has held that Plum is not responsible in Cases 1 – 3, it follows relatively easily that he is equally not responsible in Case 4; there appears, at least, to be no material difference between Cases 3 and 4 upon which responsibility might turn. Responses to Pereboom's manipulation argument generally take one of two forms: a "hard-line" reply denies the claim that Plum is not morally responsible to begin with in Case 1; whilst a "soft-line" reply denies the claim that no relevant difference can be found between two adjacent cases upon which responsibility may depend.[46]

---

[45] *Ibid.,* 79.

[46] See Michael McKenna, 'A hard-line reply to Pereboom's four-case manipulation argument' (2008a) 77(1) *Philosophy and Phenomenological Research* 142, 143; See also Ishtiyaque Haji and Stefaan E. Cuypers,

#### *13.2.3.1. The Hard-Line Reply*

Following the discussion regarding persuasion, manipulation, coercion and compulsion in section 13.2.1 of this chapter, above, it is not the causality of the intervention which distinguishes the effects of persuasion from coercion on an individual's responsibility; each are necessary and sufficient causes to change a person's decision from *not x* to *x*. Rather, it is the manner of the intervention that is relevant and, specifically, whether or not it overwhelms one of the crucial capacities identified for responsibility. Thus, it is first necessary to determine the nature of the neuroscientists' intervention in Case 1. Readily, this is some form of manipulation; the neuroscientists are neither explicitly persuading Plum with reasons to kill White, nor threatening him with severe consequences for not so killing her, nor causing any physical or psychological compulsion for him to so kill. Of the three types of manipulation considered, the neuroscientists' intervention is covert, but not deception in the sense of lying, misleading, or providing false arguments to Plum. Similarly, the intervention does not consist of pressure, such as through non-coercive threats or wearing Plum down to act in a particular way. Readily, again, the neuroscientists are exploiting Plum; specifically, they are taking the egoistic reasoning trait that he already possesses and regularly uses (*i.e.,* a "weakness") and ensuring its operation during Plum's specific deliberation on killing White.

If the neuroscientists' manipulation is exploitative, the central question becomes whether or not that exploitation overwhelms or undermines Plum's capacity to appreciate the nature and consequences of his actions. Although Pereboom does not address this capacity specifically, he does provide that Plum remains responsive to reasons and can still regulate his behaviour according to moral reasons, especially when competing egoistic reasons are weak. A charitable reading might therefore infer that Plum also remains able to appreciate the nature of his own actions (contrasting them with more moral alternatives, for example), and understands the physical, legal and moral consequences of his act of killing White. Certainly, if Plum possessed this capacity to begin with, it does not seem obvious that the fact that he was induced to deploy highly egoistic reasoning in a given circumstances would undermine his ability to still appreciate

'Hard- and soft-line responses to Pereboom's four-case manipulation argument' (2006) 21(4) *Acta Analytica* 19.

the nature of his own actions and whether, for example, killing White would have the consequence of satisfying any of his egoistic desires in the first place.

In Case 1, therefore, it is arguable that Plum is in fact responsible for killing White; notwithstanding the neuroscientists' intervention, when making his decision Plum remained capable of recognising and applying good reasons to his decision-making process, capable of controlling his bodily actions to conform with his intentions, and capable of appreciating that his actions would bring about White's death, and that this would be a legal / moral wrong. Thus, Plum's deciding mind possessed everything necessary to appreciate that killing White was wrong, to consider the various good reasons for not so killing, and to control his body to enact whichever was the final decision. Is it, then, truly counterintuitive to hold Plum responsible for killing White in Case 1?

Suppose that Plum has a twin, Peach, who is identical to Plum in all of the important ways highlighted by Pereboom in his opening conditions; the only thing that differs between Plum and Peach is the particular intervention of the neuroscientists. Thus, Peach also has first and second-order desires to kill White, is generally egoistic in his reasoning and often highly so, is responsive to reasons and capable of applying moral reasoning to his decisions, and does not act under irresistible impulse, *etc.*

> *Case 1a: A "false friend" has to ability to manipulate Peach's neural states at any time; they do not possess any special technology, but they have known Peach for such a long time that they know how to trigger different moods or states of reasoning in him. Perhaps the false friend knows that reminding Peach of his favourite egoistic character from a film or some related nickname will instil a strong sense of egoistic reasoning that Peach is prone to using. Whatever the priming method, the false friend "presses" the right metaphorical "buttons" just before Peach begins to reason about his situation, which they know will produce in him a neural state that realizes a strongly egoistic reasoning process, and which they know will deterministically result in his decision to kill White. Peach would not have killed White had the false friend not intervened, since his*

*reasoning would then not have been sufficiently egoistic to produce this decision. But at the same time, Peach's effective first-order desire to kill White conforms to his second-order desires. In addition, his process of deliberation from which the decision results is reasons-responsive; in particular, this type of process would have resulted in Peach's refraining from deciding to kill White in certain situations in which his reasons were different. His reasoning is consistent with his character because it is frequently egoistic and sometimes strongly so. Still, it is not in general exclusively egoistic, because he sometimes successfully regulates his behaviour by moral reasons, especially when the egoistic reasons are relatively weak. Peach is also not constrained to act as he does, for he does not act because of an irresistible desire – the false friend does not induce a desire of this sort.*

It is submitted that it is not obvious that Peach should be excused in this case. Many people in society may have egoistic reasoning processes, some of which are especially strong. And yet, provided that such regularly egoistic reasoning is not so compulsive as to be a malady – for example, provided that it does not follow from some sociopathy or other psychiatric disorder which undermines their capacities to otherwise respond to good reason, exhibit ordinary self-control and appreciate the nature and consequences of their action – then such egoistic people are nonetheless reasonably expected to conform their actions to the law and morals of society. Even the egoistic, *in possession of the three crucial capacities,* can recognise legal and moral rules for not acting egoistically, and apply these rules as sufficiently good reasons not to so act in particular circumstances. It does not seem intuitively correct that we would allow Peach to plead, "it was not my fault that I killed White, because a conversation with my (false) friend caused me to reason in an especially egoistic manner on this occasion." It is even more obvious that Peach is responsible if the manipulation is not intentional:

*Case 1b: A friend meets Peach and suggests that they watch a film together featuring their favourite egoistic character. Peach agrees and, seeing his favourite role model, the experience primes within him a strongly egoistic*

> *reasoning process; although his friend might have known this could happen had he considered it (because Peach's reasoning is often egoistic after all), in the event the thought did not cross the friend's mind, and he had not proposed the film viewing with any cognisance, let alone intention, of so triggering a process of egoistic reasoning in Peach's mind. Nevertheless, it just so happens that right before Peach begins to reason about his situation vis-à-vis White, watching the film "presses" the right metaphorical "buttons" to produce in him a neural state that realizes a strongly egoistic reasoning process, which deterministically results in his decision to kill White. Peach would not have killed White had the friend not intervened, since his reasoning would then not have been sufficiently egoistic…*

There is nothing in Case 1b to exculpate Peach; what is more, whilst the difference between the intentions of the friend and the false friend may equally differentiate *their* own responsibility for intervening, it is submitted that this has little impact on the responsibility of Peach in Cases 1a and 1b. Peach is obviously responsible for killing White in Case 1b; and it is readily more arguable than not that Peach remains responsible in Case 1a. The manipulation does not overwhelm or undermine any of his crucial capacities for responsibility and, notwithstanding having his strongly egoistic reasoning process exploited by the false friend in Case 1a, Peach is not readily absolved for the act of murder in the event. Even the most egoistic person is expected to recognise that their egoism does not provide justification for an act as grave and condemnable as murder. As Peach in Case 1a is identical to Plum in Case 1 in all material ways but for the specific manner of implementing the exploitative manipulation, it may be concluded that Plum is responsible for killing White also. If Plum is responsible in Case 1 then it is even easier to appreciate his responsibility in Case 2 and thereafter, in which case the manipulation argument has failed.

#### *13.2.3.2. The Soft-Line Reply*

Suppose it is allowed that the neuroscientists' intervention in Case 1 is sufficiently manipulative, or even coercive, that Plum is excused of responsibility for killing White. The soft-line reply argues that a meaningful difference can nonetheless be identified between Case 1 and Case 2 (or Cases 1 and 2, and Case 3) upon which responsibility may fairly and rationally be attributed.[47]

In Case 1 (and Cases 1a and 1b, above), it is incumbent to investigate the nature of the interventions of the neuroscientists, false friend and friend because, *in the specific circumstances of each case*, the intervention is both necessary and sufficient to procuring that Plum or Peach kill White. That is to say, the intervention (whether it is manipulative (Case 1a) or innocent (Case 1b)) was in each case necessary for procuring the subsequent act; Pereboom specifically provides in Case 1 that Plum would not act *but for the intervention*, and this specification is repeated in Cases 1a and 1b. Further, the decision to kill White in either case is in fact comprised of countless causes, most of which are unspecified. Obviously, everything must have happened to cause Plum and Peach to be alive; the combination of their genetics and upbringing must have caused their first and second-order desires to kill White, their generally, and often strongly, egoistic reasoning processes, *etc.* – all of the compatibilist conditions permitted by Pereboom; and the relevant opportunity must have arisen such that, upon deciding to kill White, Plum and Peach could in fact proceed to do so. Even here, all the relevant causes leading to any single decision are not completely indicated.

The broader point follows, at the time that the interventions occurred in Cases 1, 1a and 1b, the complete set of jointly sufficient conditions had not yet arranged to cause either Plum or Peach to kill White. In fact, this set of jointly sufficient conditions was *almost* complete, save for the addition of the neuroscientists, false friend or friend's intervention. Thus, *in the specific circumstances as they existed at the time of the intervention*, that intervention was also, in itself, sufficient to procure that the decision to kill White would absolutely follow. *Given that the remainder of the set of sufficient causes had already*

[47] See also Kristin Demetriou, 'The soft-line solution to Pereboom's Four-Case Argument' (2010) 88(4) *Australasian Journal of Philosophy* 595.

*coalesced,* the intervention was sufficient to procure the result of killing White in that moment by completing the set of sufficient causes for the decision to kill White, thus guaranteeing that outcome. Therefore, again, given the conditions as they existed as at the time of the intervention, the intervention was *both* necessary and sufficient to cause the decision to kill White. It is for this reason that we proceed to investigate the *nature* of the intervention, *i.e.,* is it persuasive, manipulative, coercive or compulsive, as this determination is relevant to the question of responsibility.

The same cannot be said for Case 2; indeed, in the specific circumstances of Case 2 *as they existed at the time of the intervention*, that intervention was neither necessary nor sufficient to secure that Plum will kill White 40 years later, even though it may nonetheless deterministically cause that he does. The intervention may nonetheless be *a* cause of Plum's behaviour but, *in the moment of the intervention*, it is no longer a necessary or sufficient cause of the same; for this reason, the intervention (however it is categorised) does not meet the requirements for exculpatory manipulation, coercion or compulsion. In Case 2, the neuroscientists program Plum at the beginning of life with the often (but not always) strongly egoistic reasoning process; yet, whilst this is the same reasoning process they caused him to have in Case 1, in that case they specifically caused the egoistic reasoning process to be engaged *as Plum was deciding whether or not to kill White*.

The same intervention in Case 2 produces no such guarantee – that is to say, just because Plum has been programmed with an often strongly egoistic reasoning process does not secure that that reasoning process will be in operation 40 years later when he deliberates about killing White. That guarantee was available in Case 1 *precisely because* the neuroscientists intervened contemporaneously with Plum's deliberation, securing that his strongly egoistic reasoning process was in operation at that time of, and in relation to, the particular decision to kill White. It was specifically Plum's *decision to act* that the intervention targeted and affected.

It is clear that the intervention in Case 2 is no longer a necessary condition to securing the decision to Kill White by again considering Peach, Plum's twin. It must be recalled

that the neuroscientists intervention is simply to cause a *strongly* egoistic reasoning process, whereas Plum is described in the precursory material as already being often egoistic in his reasoning, such that the neuroscientists' intervention does not create conditions that are outside of Plum's normal character.

> *Case 2a: Peach is identical to Plum except that he was not programmed by a team of neuroscientists; rather, at some point during his upbringing Peach read an influential book about successful reasoning strategies, resulting in his own reasoning being often but not always egoistic (as in Case 1), and at times strongly so, with the consequence that in his current circumstances he is causally determined to engage in the egoistic reasons-responsive process of deliberation and to have the set of first and second-order desires that result in his decision to kill White. Peach has the general ability to regulate his actions by moral reasons, but in his circumstances, due to the strongly egoistic nature of his deliberative reasoning, he is causally determined to make his decision to kill. Yet he does not decide as he does because of an irresistible desire. The neural realization of his reasoning process and of his decision is exactly the same as it is in Case 1 (although their causal histories are different).*

What Case 2a demonstrates is that there are numerous plausible routes by which a person like Peach (or Plum), already instilled with the egoistic characteristics and desires to kill White, may in fact come about to so killing. The intervention of the neuroscientists was necessary in Case 1 when, *at the time of their intervention,* Plum had not yet decided to kill White and would not have so decided without the intervention. In Case 2, as shown by way of Case 2a, any number of factors or events may operate as contributory causes to a later decision some 40 years hence. In Case 2, *at the time of their intervention*, it cannot be said that the neuroscientists' intervention was a necessary condition for Plum killing White as, in the 40 years which passed before the decision itself, it is possible and perfectly plausible that other causes could have led a person like Plum to the same result, just as they did with Peach in Case 2a.

The intervention was a necessary condition to securing that Plum kill White in Case 1 because he would not otherwise have done so; *at the time of the intervention, such that it was immediately prior to the decision to act*, it was necessary for the neuroscientists to intervene to ensure that Plum was sufficiently egoistic. The intervention of the neuroscientists was not a necessary condition in Case 2 because, had they not intervened *at the time that they did*, allowing for the passage of 40 years from the time of their non-intervention to Plum's decision to act, that decision could plausibly and readily have been procured by any number of different causal routes, such as demonstrated by Peach in Case 2a.

The intervention of the neuroscientists in Case 2 is also not sufficient *at that time*, as can be seen from Case 2b.

> *Case 2b: Peach is just like Plum, including having been programmed by a team of neuroscientists at the beginning of his life so that his reasoning is often but not always egoistic (as in Case 1), and at times strongly so. Peach has the set of first and second-order desires to kill White, along with the general ability to regulate his actions by moral reasons. During his early adulthood, noticing the immorality of his own egoistic reasoning and its consequences throughout his youth, over time Plum learned to reason more carefully and always search for any overriding moral reasons not to act when he noticed that he was about to make a particularly egoistic decision. Consequently, in his circumstances, due to having learned to catch himself from making decisions that are too egoistic or immoral, he is causally determined by the neuroscientists' original intervention to decide not to kill White.*

When the neuroscientists acted in Case 1, their intervention was sufficient because *at that particular time* it completed a broader set of sufficient conditions and secured Plum's action. *At the time of the intervention* in Case 2, however, the broader set of sufficient conditions required to guarantee that Plum kills White 40 years hence have not yet coalesced, including most obviously Plum's reaching adulthood and obtaining the

opportunity to kill White. *At the time of the intervention* in Case 2, the intervention does not guarantee that it follows that Plum will kill White 40 years later. Not only may any number of casual events intervene over the following decades, but *the fact that the neuroscientists intervened earlier rather than later* could plausibly cause the opposite of their intentions, as in Case 2b. To propose that the neuroscientists' intervention in Case 2 is qualitatively identical to that in Case 1 would require that the intervention in Case 2 had no possible interaction or impact whatsoever on Plum's subsequent development for the following four decades, until the precise moment that he came to consider killing White and had the opportunity to do so.

Intervening at the point of a particular decision to act (as in Case 1) can be sufficient to secure a given outcome, such as a when any decision is persuaded, manipulated, coerced or compelled. When a decision to act is made, all of the necessary and sufficient causes have coalesced to result in that decision; moments before during the process of deliberation, therefore, it may be appreciated that the requisite set of sufficient causes has *almost* coalesced, in which circumstances a single intervention can itself be sufficient to complete that set of causes and push the particular decision over the threshold in a given direction.

However, intervening in a decision to act some 40 years before it even arises is quite different. A far fewer number of the set of sufficient causes of that later decision have yet coalesced, such that the particular intervention cannot be regarded as itself sufficient *at that time*; it becomes one of any number of causes within an *incomplete set of sufficient causes*, but is not individually sufficient in the particular circumstances as they exist at the time of the intervention. Moreover, not only does the time following that intervention allow for any manner of other causes to also intervene and impact upon the later decision, but that earlier intervention itself could plausibly cause the very reverse of what the neuroscientists intended. Each of these points alongside Plum's responsibility in Case 2 may be demonstrated through a final counterexample following a (somewhat) more real-world narrative.

*Case 2c: White wants to have a baby and, for whatever reasons, procures one through in-vitro fertilisation. Having sequenced the human genome incredibly precisely, the fertility doctors have the choice of implanting one of three embryos. Each are identical save for one trait: Scarlet's embryo has the psychopath gene which, for the sake of argument, always results in matricide; Plum's embryo has the often-strongly egoistic gene; and Peach's embryo has the pacifist gene. In the event, each child would grow up in an environment where they are encouraged to be often egoistic (but not always) in their reasoning, are taught how to apply moral reasons to their decisions and become responsive to reasons etc. and, unfortunately, would each come to have first and second-order desires to kill White, their mother. The fertility doctors select Plum for implantation who, in his current circumstances, is causally determined to engage in the egoistic reasons-responsive process of deliberation and to have the set of first and second-order desires that result in his decision to kill White. Plum has the general ability to regulate his actions by moral reasons, but in his circumstances, due to the strongly egoistic nature of his deliberative reasoning, he is causally determined to make his decision to kill. Yet he does not decide as he does because of an irresistible desire. The neural realization of his reasoning process and of his decision is exactly the same as it is in Case 1 (although their causal histories are different).*

First, the fertility doctor's intervention in choosing Plum is clearly not a necessary condition for the subsequent decision to kill White; they could have chosen Scarlet instead who, for the sake of argument, would certainly commit matricide as a result of her psychopathy. Second, it cannot be said in any reasonable sense that selecting Plum was a sufficient cause of White's death; for the reasons given previously, his selection at birth offered no surety that his often (*but not always*) egoistic reasoning would be in play 40 years later when he killed White. Plum's genetically incurred egoistic reasoning process may be part of a set of together sufficient conditions which fully coalesce later to procure his decision to kill White, *but at the time of the intervention* causing Plum to be often egoistic was not sufficient *in that moment* to guarantee his later decision to kill.

And third, there is no intuitive sense in which Plum is absolved of responsibility in Case 2c, even though he was deterministically caused to have a trait which subsequently caused him to kill White. Even if the fertility doctors were nefarious and selected Plum because they wanted him to kill White, and were therefore exploiting the existence of his genetic trait to procure this end, this manipulation would hardly be counted as sufficient to absolve responsibility. If Plum nevertheless retains the crucial capacities discussed, then the existence or otherwise of certain character traits neither proves nor disproves responsibility, howsoever those traits were themselves caused. The only meaningful differences between Case 2 and 2c is that fertility doctors intervene as opposed to neuroscientists, and they cause Plum to have a strongly egoistic reasoning process by selecting it within his genetic makeup rather than using some form of neural programming. Therefore, even if Plum is deemed not to be responsible in Case 1, there are strong reasons why he is responsible in Case 2 and thereafter, and the manipulation argument fails.

Pereboom specifically provides that the neuroscientists' intervention is a necessary cause of Plum's decision to kill in Case 1, but not in Case 2. This may be an oversight or, more charitably, it could be assumed that this necessity carries over, such that Plum would not decide to kill White but for the neuroscientists' intervention in Case 2 also. Similarly, whereas it is clear that the neuroscientists' intervention in Case 1 was sufficient in the circumstances to ensure that Plum decides a particular way, the same intervention deployed some 40 years prior in Case 2 is clearly not sufficient *in those circumstances* to secure the same decision. Many more causes need to coalesce over the course of Plum's coming life before he reaches the choice of whether or not to kill White, and the neuroscientists would need to take considerably greater and more intrusive action in order to ensure the coalescence of those causes to meet their intentions. Thus, contrary to Pereboom's argument, the same action can have a different causative nature when deployed in different circumstances and at different times.

In Case 1, the intervention necessarily and sufficiently affected a *decision to act* to which responsibility could attribute, and was thus in the nature of a potentially excusatory manipulation. In Case 2, however, the intervention was (in the circumstances as they existed at that time) a mere, but neither necessary nor sufficient, cause of Plum subsequent

behaviour. The intervention did not, therefore, meet the criteria of responsibility-absolving manipulation or coercion, notwithstanding that it may nevertheless have contributed to Plum's decision. Were Pereboom to suggest that the same intervention by the neuroscientists in Case 2 was both necessary and sufficient to secure Plum's subsequent decision 40 years hence, would be to suggest both that no other possible causal route to White's death exists, and that no possible intervening cause could prevent White's death from occurring after the intervention. This would no longer be an argument from causal determinism, but one from fatalism.

Pereboom argues that it is 'unprincipled to claim that [in Case 2], by contrast with Case 1, Plum is morally responsible because the length of time between the programming and his decision is now great enough. Whether the programming occurs a few seconds before or forty years prior to the action seem irrelevant to the question of his moral responsibility.'[48] This is not so, however, because the passage of time can change the causative nature of the same intervention – pouring a glass of water into a half-filled bucket is not, *in that moment*, necessary or sufficient to cause the bucket to overflow; but performing the same action of pouring a glass of water into a bucket filled to the brim is, *in that moment*, necessary and sufficient to cause the bucket to overflow. In Case 1, the neuroscientists are assured that their intervention is going to affect *the specific decision they wish to influence* and cause a particular outcome in that moment because they are specifically targeting Plum's decision to act, the very thing upon which responsibility may or may not be attached. In Case 2, the same intervention does not, *at the time that it is made,* guarantee that the decision whether or not to kill White will even arise in the future, and provides even less guarantee that Plum's programmed egoistic reasoning will be sufficiently strong at the time of that decision to secure a particular outcome.

Thus, the intervention is not even sufficiently targeted at a specific *decision to act* upon which responsibility may attach. What is more, if the neuroscientists really wanted to assure Plum's future killing of White 40 years later, aside from orchestrating the relative motive and opportunity for Plum's decision to kill to arise (which would surely involve considerably greater overall manipulation), they would surely need to program a

[48] Pereboom (2014), 77 – 78.

significantly stronger form of "reasoning", motive or desire in order to ensure that it was engaged as and when the decision to kill White actually arose four decades later. However, this would be approaching upon some kind of "irresistible" impulse the Pereboom denies, or lack of reasons-responsiveness that he strives to maintain. That which the neuroscientists would *actually* need to do in Case 2 in order to necessarily and sufficiently cause Plum to kill White would be so considerably more intrusive, manipulative or even psychologically coercive that Plum's responsibility would be affected. If, however, the neuroscientists merely act in Case 2 as they did in Case 1, but at an earlier time, then it is appreciable why Plum can be responsible in Case 2 but not in Case 1, because the causative nature of the intervention has changed.

Mele offers a similar critique of Case 2 of the manipulation argument, highlighting how the same intervention of causing him to be sometimes (but not always) strongly egoistic could have resulted in different effects:

> 'Normal agents learn how to weigh reasons for action. For example, a young agent who weighs reasons very egoistically may suffer as a consequence and learn that things go better for him when he weighs the interests of others more heavily as reasons. His deliberative style might gradually become significantly less egoistic, and, along the way, his less egoistic actions might have reinforcing consequences that help to produce in him increased concern for the welfare of those around him. This increased concern would presumably have an effect on his evolving deliberative habits. The story of the normal evolution of a particular agent's deliberative style is a long one. The point here is that in the 2 series Plum is cut off from such evolution regarding his procedure for weighing reasons.'[49]

Here, Mele has offered a scenario not too dissimilar from Case 2b, above, in which the intervention perversely causes behaviour opposite to that intended by the manipulator. The broader point is that, whereas the manipulation in Case 1 directly impacts upon

[49] Alfred R. Mele, *Free Will and Luck* (Oxford University Press 2006), 142 – 143.

Plum's decision to act in a plausibly acceptable manner, for the same manipulation to *guarantee* the same effect in Case 2 would be to remove Plum from all possible alternative causal paths, which is no longer a plausibly acceptable supposition. An alternate but related argument follows that Plum in Case 1 has arrived at the moment of deliberation without any intervention, whereas Plum in Case 2 is fundamentally different *because of* the intervention that occurred at his birth; that same intervention at an early time has changed the causal chain and cannot therefore guarantee the same result without fatalistically denying all other possible causal paths that could follow from the intervention to the moment of the decision to kill White.

#### *13.2.3.3. Empirical Research on the Four-Case Manipulation Argument*

Pereboom's Four-Case Manipulation Argument appeals to the intuitions that the reader has towards Plum's moral responsibility in each case. Feltz reports an illuminating empirical experiment where variants of the Four-Case Argument designed to be more readily understandable (and appealing less to esoteric philosophical terms) were administered to subjects in order to gather their actual intuitive responses to each case.[50] In the first experiment, 112 subjects were randomly assigned to respond to all four cases of the argument or a single case only, after which they responded to various statements regarding Plum's responsibility, blameworthiness and deservedness of punishment on a seven-point Likert scale. On average, subjects found Plum to be morally responsible in Cases 2 – 4, only finding him not to be responsible in the case of *intentional direct manipulation* of the particular decision in Case 1 (see ***figure m.***).[51] These findings support the soft-line response, identifying a relevant difference between Cases 1 and 2 upon which Plum is responsible for his actions in the latter case.

In experiment 2, the cases were varied such that the manipulation in Cases 1 and 2 was non-intentional, analogous to Case 1b above. A total of 197 subjects were again assigned to respond to all four cases or a single case, differing in the fact of the non-intentional nature of the manipulation. Here, subjects judged on average that Plum was responsible

[50] Adam Feltz, 'Pereboom and premises: Asking the right questions in the experimental philosophy of free will' (2013) 22(1) *Consciousness and Cognition* 53.
[51] *Ibid.,* 57 & 59.

in all four cases, supporting the hard-line response to Pereboom's argument (see ***figure m.***). Thus, there is tentative evidence that Pereboom's intuitions about Plum's moral responsibility do not necessarily hold true for the public in general; whether defeated by a hard- or soft-line response, this hard-incompatibilist argument against moral responsibility does not appear to hold up to scrutiny.

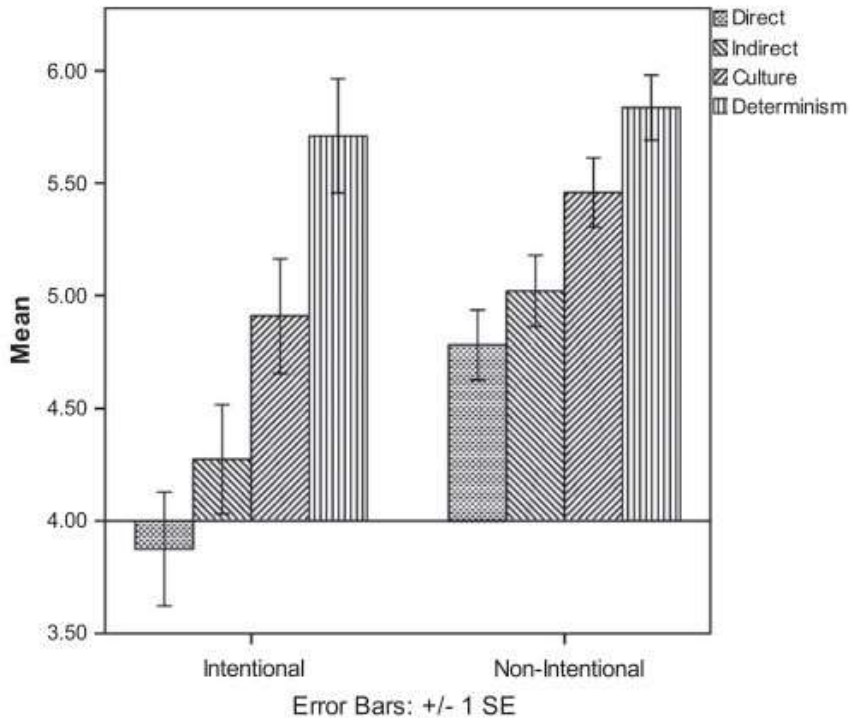


***Fig. m – Mean responses to the four-case argument with intentional and non-intentional manipulator.***

### *13.2.4. A Note on Indeterminism*

The present thesis has been undertaken with the underlying assumption of a deterministic universe within and because of which metaphysical free will is precluded. One counterargument that is made suggests that the potential for free will could emerge from indeterministic features of the universe. The argument appeals in particular to the possible influence of quantum mechanics which, on the one hand, is premised upon the fundamental indeterminacy of the sub-atomic universe whilst, on the other hand, is traditionally confined to that sub-atomic universe such that the indeterministic principles of quantum mechanics are not presumed to have macroscopic effects, and the classical

physics of the macroscopic universe is presumed to remain causally deterministic.[52] This was especially presumed to be the case once considering the level of biological systems; quantum effects are sub-atomic, delicate, and easily disturbed or destroyed through interaction with surroundings and would not, therefore be expected to produce macroscopic effects in the 'warm, wet and noisy' world of biological systems.[53]

This latter assumption has been challenged by the recently emerging scientific study of quantum biology, which explains various biological phenomena through the operation of key underlying features of quantum mechanics such as the uncertainty principle, wave-particle duality, superposition and quantum tunnelling.[54] For example, the sense of smell is traditionally explained through a "lock-and-key" model whereby molecules of certain shapes fit into specific nasal receptors like a key in a lock which, when correctly matched, triggers signals to the brain to produce the sensation of a particular smell. However, this theory fell under criticism from various directions, such as whether around 400 different smell receptors could account for the hundreds-of-thousands of different recognisable smells, or why certain chemicals would smell similar whilst having different molecular shapes or, *vice versa,* might smell different despite having similar molecular shapes.

Prior to the lock-and-key model even being developed, Dyson proposed in 1928 that the brain might interpret the characteristic frequencies at which different molecules vibrate as their different smells, drawing an analogy to the way in which the brain interprets different vibrational frequencies of light as colour;[55] however, the theory failed to gain early traction. The idea was subsequently revived by Turin in 1996, who proposed a more specific mechanism by which such a "vibrational" sense of smell might operate, namely through quantum tunnelling.[56] In brief, the theory proposes that when an odorous

[52] See generally David Hodgson, 'Quantum physics, consciousness, and free will' in Kane R. (ed.), *The Oxford Handbook of Free Will* (Oxford University Press 2011), 57 – 59.
[53] Stuart R. Hameroff, 'Biological feasibility of quantum approaches to consciousness' in van Loocke P. (ed.), *The Physical Nature of Consciousness* (John Benjamins Publishing Company 2001), 21.
[54] Graham R. Fleming and Gregory D. Scholes, 'Quantum biology: introduction' in Mohseni M., Omar Y., Engel G. and Plenio M. B. (eds.), *Quantum Effects in Biology* (Cambridge University Press 2014), 3 – 4.
[55] Malcolm Dyson, 'Some aspects of the vibration theory of odor' (1928) 19 *Perfumery and Essential Oil Record* 456.
[56] Luca Turin, 'A spectroscopic mechanism for primary olfactory reception' (1996) 21(6) *Chemical Senses* 773.

molecule meets a nasal receptor, electrons can "burrow" through the energy barrier to unleash signals to the brain from the other side, *if* the quantised energy level of the electron matches the vibrational frequency of the odorant molecule. Thus, it is the quantum vibration of molecules at the atomic level which result in the macroscopic phenomenon of smell.

A subsequent paper by Brookes, Hartoutshiou, Horsfield and Stoneham[57] modelled the mechanisms involved to determine whether they were viable in principle. Setting parameters appropriate to biomolecular systems, they showed that the proposed "swipe card model" of smell was 'consistent both with the underlying physics and with observed features of smell', such as the fact that molecules with similar shapes can elicit different smell responses due to, it is proposed, their different vibrational frequencies. In 2011, Turin *et. al.* published experimental evidence strongly supporting the theory;[58] they demonstrated that fruit flies could distinguish between two forms of acetophenone differing between hydrogen and deuterium in their composition. Whereas the compounds have the same molecular shape, they vibrate at different frequencies; thus, the successful differentiation by the fruit flies could only be explained by reference to quantum effects occurring at the macroscopic level. Evidence for quantum effects in other biological processes is increasingly forthcoming, including in relation to photosynthesis, avian navigation using the Earth's magnetic field, and bioluminescence, to name some examples.[59] It is even hypothesised that quantum effects may have been instrumental in either or both of the genesis of life itself and / or the random mutations in DNA which drive evolution.[60]

[57] Jennifer C. Brookes, FIlio Hartoutsiou, A. P. Horsfield and A. M. Stoneham, 'Could humans recognise odor by phonon assisted tunnelling?' (2007) 98(3) *Physical Review Letters* 3.
[58] Maria Isabel Franco, Luca Turin, Andreas Mershin and Efthimios M. C. Skoulakis, 'Molecular vibration-sensing component in *Drosophila melanogaster* olfaction' (2011) 108(9) *Proceedings of the National Academy of Sciences* 3797.
[59] Paola Lecca and Angela Re, *Theoretical Physics for Biological Systems* (CRC Press 2019), 14 – 19; see also Mohseni M., Omar Y., Engel G. and Plenio M. B. (eds.), *Quantum Effects in Biology* (Cambridge University Press 2014).
[60] Lars Jaeger, *The Second Quantum Revolution: From Entanglement to Quantum Computing and Other Super-Technologies* (Springer Nature Switzerland 2018), 261; see also Johnjoe McFadden and Jim Al-Khalili, *Life on the Edge: The Coming of Age of Quantum Biology* (Crown Publishers 2014).

With growing and persuasive evidence that quantum phenomena can pertain effects in the warm, wet and noisy macroscopic world of biology, it begs the question whether such quantum effects might also be at play in the human brain. For example, a quantum theory of consciousness has been controversially defended by Penrose and Hameroff.[61] A full examination of all such proposals is neither warranted nor necessary; however, two modes by which it is speculated that quantum effects could impact upon decisions to act may be considered briefly, as it is these specific decisions which relate to the concept of responsibility under investigation throughout the present thesis.

Quantum fluctuation refers to the phenomenon whereby opposing pairs of 'charged particles like electrons and positrons are constantly being created and destroyed.'[62] The phenomenon emerges from Heisenberg's uncertainty principle which states that no particle can ever have an exact position and momentum in space and time, but exists as a probability function in all possible positions until collapse of the wave function. Thus, quantum fluctuation is the 'temporary emergence of energetic particles from nothing, as allowed by the uncertainty principle.'[63] Compelling experimental evidence for quantum fluctuation is traditionally drawn from the "Casimir Effect";[64] more recently, researchers at the Laser Interferometer Gravitational-wave Observatory ('LIGO') have experimentally confirmed the influence of quantum fluctuations on the motion of large, human-scale macroscopic objects, namely the mirrors that are fundamental to the LIGO.[65]

Jedlicka contends that, contrary to the orthodox view that quantum fluctuations would be self-averaging in the warm, wet and noisy environment of the human brain, 'because of

[61] Stuart R. Hameroff and Roger Penrose, 'Orchestrated reduction of quantum coherence in brin microtubules: A model for consciousness' in Hameroff S. R., Kaszniak A. W. and Scott A. C. (eds.), *Toward a Science of Consciousness: The First Tucson Discussions and Debates – Vol. I* (Massachusetts Institute of Technology Press 1996); Stuart R. Hameroff and Roger Penrose, 'Consciousness in the universe: A review of the "Orch OR" theory' (2014) 11(1) *Physics of Life Reviews* 39.
[62] B. G. Sidharth, *The Universe of Fluctuations: The Architecture of Spacetime and the Universe* (Springer 2005), 77.
[63] Luciano Boi, *The Quantum Vacuum: A Scientific and Philosophical Concept, from Electrodynamics to String Theory and the Geometry of the Microscopic World* (Johns Hopkins University Press 2011), 1.
[64] Andrei A. Bytsenko, G. Cognola, E. Elizalde, V. Moretti and S. Zerbini, *Analytic Aspects of Quantum Fields* (World Scientific Publishing Co. 2003), Ch. 9.
[65] Haocun Yu, L. McCuller, M. Tse, N. Kijbunchoo, L. Barsotti, N. Mavalvala and members of the LIGO Scientific Collaboration, 'Quantum correlations between lights and the kilogram-mass mirrors of LIGO' (2020) 583(7814) *Nature* 43.

an extreme sensitivity to initial conditions, in complex [non-linear] systems the microscopic fluctuations may be amplified and thereby affect the system's behavior.'[66] Thus, the hypothesis argues, where part of the operation of neurons fundamentally relies on temporary shifts from negative-to-positive action potentials across the neuron's membrane, which is itself caused by the opening and closing of ion gates along that membrane, the sudden emergence of electrically charged particles (*i.e.,* electrons and positrons) in sufficient quantities and in the correct locations could have resultant effects on neuronal action potentials and / or ion gating. In this regard, for example, Vaziri and Plenio[67] suggest that quantum physics is likely to influence certain stochastic events in the brain such as the opening of ion channels. Meanwhile, Glimcher writes:

> '[D]ata suggest that membrane voltage is the product of interactions at the atomic level, many of which are governed by quantum physics and thus are truly indeterminate events. Because of the tiny scale at which these processes operate, interactions between action potentials and transmitter release as well as interactions between transmitter molecules and post-synaptic receptors may be, and indeed seem likely to be, fundamentally indeterminate.'[68]

In a similar vein, McFadden writes:

> 'If neurones poised on the dynamics of individual membrane proteins are critical to the initiation of a particular course of motor action or cognitive process, then the consequent action or cognitive processes will be subject to non-deterministic quantum dynamics.'[69]

---

[66] Peter Jedlicka, 'Revisiting the quantum brain hypothesis: Toward quantum (neuro)biology?' (2017) 10 *Frontiers in Molecular Neuroscience* 1.
[67] Alipasha Vaziri and Martin B. Plenio, 'Quantum coherence in ion channels: resonances, transport and verification' (2010) 12(8) *New Journal of Physics* 085001.
[68] Paul W. Glimcher, 'Indeterminacy in brain and behavior' (2005) 56(1) *Annual Review of Psychology* 25, 49.
[69] Johnjoe McFadden, 'The conscious electromagnetic information (cemi) field theory' (2002) 9(8) *Journal of Consciousness Studies* 45.

Jedlicka argues further that, even if quantum effects could not be observed directly at the macroscopic level, they could still indirectly influence the functions of highly non-linear systems such as the brain and broader nervous system, which consists of a nested hierarchy of complex non-linear networks. In such conditions, small quantum fluctuations may not be averaged out but, in fact, become amplified across iterative hierarchies with non-liner dynamics; 'quantum fluctuations on the lowest level of scale may influence the initial state of the next level of scale, while the higher levels shape the boundary conditions of the lower ones.'[70] Rather than cancel out quantum effects, Satinover argues that they would be exploited:

> 'Quantum dynamics alters the final outcomes of computation at all levels – not by producing classically impossible solutions but by having a profound effect on which of many possible solutions are actually selected.'[71]

Similarly, Sompolinsky writes that 'chaos within the brain may amplify enormously the small quantum fluctuations… to a degree that will affect the timing of spikes in neurons.'[72] Even Koch, a critic of quantum effects in the brain, acknowledges:

> 'What cannot be ruled out is that tiny quantum fluctuations deep in the brain are amplified by deterministic chaos and will ultimately lead to behavioral choices.'[73]

How might appear the effects of quantum fluctuation on a brain that is deciding to act? It is hypothesised that these effects might be characterised as providing the "spark" or "nudge" to any given decision; "spark" refers to contributing to the "spark of inspiration"

[70] Jedlicka (2017), 4.
[71] Jeffrey Satinover, *The Quantum Brain: The Search for Freedom and the Next Generation of Man* (John Wiley & Sons 2001), 210.
[72] Haim Sompolinsky, 'A scientific perspective on human choice' in Berger Y. and Shatz D. (eds.), *Judaism, Science, and Moral Responsibility* (Rowman & Littlefield Publishers 2006), 32.
[73] Christof Koch, 'Free will, physics, biology, and the brain' in Murphy N., Ellis G. F. R. and O'Connor T. (eds.), *Downward Causation and the Neurobiology of Free Will* (Springer-Verlag Berlin Heidelberg 2009), 40.

that produces a novel idea, whilst "nudge" refers to contributing to pushing a decision that has "almost been made" over the requisite threshold to become a final decision to act.

An original idea is often colloquially described as the novel combination of two previously unconnected ideas. Thus, it can be imagined how two concepts or ideas are represented by distinct networks of neurons in the brain, whilst the pathways connecting these two networks have not yet reached sufficient excitation for action potentials to be triggered. It may be hypothesised that quantum fluctuations occurring in the relevant locations of the brain along that pathway such as to impact upon the ion gates of those pathway neurons connecting two distinct networks of representations could initiate those pathways such as to provide a novel connection identified as an original idea. This would feasibly present as the "spark of inspiration"; the "eureka moment"; the point at which sudden clarity is achieved in reaching a novel solution to a given problem.

Alternatively, it may be imagined that the neuronal network representing a particular decision outcome has amounted such excitability that it has *almost* reached the threshold of becoming the definitive decision, and in which circumstances quantum fluctuations occurring in the relevant locations of the brain provide the final input of excitation which pushes that network over the threshold to become the final decision. In these circumstances, the quantum fluctuations have "nudged" one option over the finish line before another option had the opportunity to become the final decision. By these two operations, it is therefore appreciable how the indeterminacy of quantum mechanics *could* be conceived to contribute to some manner of metaphysical freedom within human decision-making, whether by providing the spark of inspiration for an agent's novel decision or by nudging a particular option over the threshold to become a final operative decision to act.

As Horst writes, however, quantum indeterminacy ultimately presents a blind alley to the discussion of free will and responsibility; for indeterminacy does not provide any greater foundation for metaphysical freedom than universal determinism:

> 'If my actions are results of probabilistic laws, with only brute chance determining everything that happens at indeterministic junctures, this is every bit as inconsistent with voluntary spontaneity as is determinism. *Acting randomly is not the same thing as acting freely*. It may be possible to freely make a random choice. But randomness grounded in brute physical chance does not amount to free will. Indeed, if my actions are ultimately governed by chance phenomena involving the quarks and leptons that make up my body, my actions are not free.'[74]

Thus, even if quantum effects occur in the brain to cause or influence certain decisions, their indeterminacy offers no route to metaphysical free will. More to the point, provided that the agent is in possession of the three crucial capacities, there is no reason why the indeterminacy involved in their arriving at a particular decision to act should absolve that individual of responsibility. Merely because one option to act has been arrived at through indeterministic rather than deterministic processes does not of itself impact at all upon the ascription of responsibility. If the agent possessed the three capacities, then they still decided to act in circumstances where they possessed everything required to appreciate that that action contravened a given legal / moral rule, which was itself a good reason not to so act. Consequently, responsibility still cannot be premised upon any notion of a metaphysically "free" agent, and must further be premised upon something other than the fact of the existence of indeterminate causal chains, just as the present thesis premises responsibility on something other than the presumed fact of causal determinism, namely the capacities of the deciding brain.

[74] Steven Horst, *Laws, Mind, and Free Will* (Massachusetts Institute of Technology Press 2011), 103 (emphasis added).

# 14. Conclusions

> 'We all were sea-swallowed, though some case again, and by that destiny to perform and act whereof what's past is prologue, what to come in yours, and my, discharge.'
>
> - William Shakespeare, 1611.[1]

The present thesis investigated the central research question, *how can people rationally be held responsible for their actions in a deterministic universe absent of metaphysically free decision-making?* The thesis began with a key underlying presumption that metaphysical free will *does not* exist, and that the (macroscopic) universe is fundamentally deterministic. For present purposes, metaphysical free will refers to either or both of the philosophical claims that: a) it is possible for a brain to make a different decision to that which it would otherwise make when faced with the same decision in identical conditions (*i.e.,* the "principle of alternative possibilities"), or that; b) it is possible for a brain to make a decision that is completely independent of any prior causes (*i.e.,* a decision that is an "original," uncaused cause, or *causa sui*).

The current approach to ascribing criminal legal responsibilty in UK law was adopted as starting model, which naturally shares a great many concepts and principles with similar approaches to legal responsibility applied in common law jurisdictions around the world. The thesis explored empirical and theoretical research from the neuropsychology of human decision-making in which, generally speaking, the *status quo* adopts a broadly deterministic perspective on decision-making and rejects notions of Cartesian dualism, eliminating the possibility for an independent, causally efficacious, and metaphysically free "mind." Conclusions were subsequently taken from the body of neuropsychological research and relevant implications were applied to the various concepts and principles

[1] William Shakespeare, *The Tempest* (Raffel B. (ed.), Yale University Press 2006), 59.

that comprise the current approach to legal responsibility. Thus, the thesis adopted a meliorative process, investigating whether and to what extent the existing approach to legal responsibility withstands scrutiny from the *status quo* of neuropsychology research, all the while maintaining a presumption against the existence of metaphysical free will.

*First Objective*

The first objective of the thesis was to elucidate and expand upon theories of decision-making from psychology and neuroscience, and relate the current state of the art to relevant aspects of legal responsibility. The neuropsychology research was explored by first disambiguating any decision to act into the five components of *what* to do, *how* to do it, *when* to do it, *whether* or not to do it, and *why* to do it. Multi-alternative decision field theory was considered as a general model of how the brain makes a decision between alternative competing options. The model proposes that populations of neurons represent the different available options under consideration. Each population competes with one another to recruit "evidence" – reflected in the increase in valence or excitability of one such representative population other the others. Such evidence includes endogenous factors such as memories regarding different decision options and the emotion or affect accompanying those memories; and exogenous factors experienced through the senses, such as primes or persuasive arguments from another person or other source.

The various representative populations of neurons "compete" to reach a threshold which may be set "naturally" or "artificially" – respectively, for example, the threshold might simply be the point at which one representative neuronal population has significantly greater valence than the others, or might consist of some external time pressure by which time a decision must be reached. The thesis subsequently proposed the integration of multi-alternative decision field theory with a distributed consensus model which proposes that neural networks represent different goals and actions on multiple levels, reaching a unified decision when a distributed consensus is reached by the representative neuronal populations across those levels. This offers a descriptive account of how each of the *what*, *how, when*, *whether*, and *why* components of any decision may be processed simultaneously in parallel by the brain across reciprocally connected levels of processing. This account may also explain how the brain manages to process each component of any

given decision so rapidly into what appears to consciousness as a single unified decision, and how input and biases from different endogenous and exogenous sources can impact across each level of a given decision.

The thesis proceeded to investigate each of the *what*, *how, when*, *whether*, and *why* components of a decision in turn, drawing implications from the conclusions of neuropsychological research that are relevant to concepts and principles in legal responsibility. As these various conclusions and their implications are concerned with how the brain operates, they are naturally most relevant to the legal concept of *mens rea*. Within the currently existing law, *mens rea* refers to the requirement that any criminal action (*actus reus*) is accompanied by a relevant guilty state of mind such as intention, recklessness and dishonesty, *etc.*, as required according to each particular criminal offence. *Mens rea* also includes a second concept of volition, requiring that any criminal act has been committed volitionally; this consists of the defendant possessing the capacity to recognise and respond to reasons in their decision-making, and to exercise conscious control over their decisions and actions. Whilst the law adopts a rebuttable presumption that all adults act volitionally unless proven otherwise, the existence of a relevant guilty state of mind at the time of committing the *actus reus* of an offence is one of the components that the prosecution must generally prove beyond reasonable doubt during a criminal trial.

The *what* component of any decision contains the very essence of a potentially criminal decision to act; a decision to steal something, attack another person, falsify a tax return, drive after drinking, or kill another person are all decisions about *what* to do. Taking some of the most salient points from the neuropsychological research, evidence showed that decisions of what to do can be reached through conscious, deliberative processes or as the result of automatic, unconscious processing. Equally, the decision of what to do may consist of some entirely endogenously selected option, or an option that has been primed in the brain by some exogenous source without any conscious awareness of the individual actor. Crucially, the legal investigation into subjective states of mind such as intention or recklessness cannot necessarily provide any distinction between the two – *i.e.,* whether any particular intention has been reached consciously and deliberately or as a result of

unconscious automatic processes, nor whether that intention reflects an individual's endogenous goals and desires or is the result of priming from some exogenous source. It was concluded that an unknown proportion of the *what* component of our decisions arise from external cues and unconscious processes over which people have little to no subjective insight or conscious control. Consequently, the legal focus on subjective states of mind likely reveals little about whether or not a person can truly be said to have consciously and deliberatively chosen a particular (criminal) action.

The *how* component of a decision concerns the way in which the brain both plans and monitors the physical actions that are necessary to convert a mere decision to act into actual bodily actions. Two particularly notable points may be summarised; first, evidence suggests a close connection between the *what* and *how* components of a decision, it being likely that the brain prepares multiple action plans (*i.e., how)* and accesses their relative merits and shortcomings *as part of the process of deciding what* to do. This is perfectly logical on the one hand – the relative ease or difficulty with which each competing plan could actually be carried out is a highly relevant factor to determining which plan to actually pursue (*i.e., what*), whilst the preparation of multiple plans of action enables people to switch more rapidly between competing options, for example, when the demands of a particular situation require responsive actions to be both fast and adaptive. On the other hand, this means that if one or more such decision alternatives under consideration consists of criminal activity, the brain is likely already in a stage of preparation to commit that criminal act before it has actually been selected as the outcome of a final decision. In such a stage of preparation, it perhaps does not require much more for even ordinarily law-abiding individuals to "tip over" into criminal conduct.

Second, the intimately connected components of *what* to do and *how* to do it also each contribute to the subjective sense of agency that people experience over their decisions to act. Specifically, evidence suggests that the sense of agency is partially constructed by prospective processes related to the selection between different decision options, and retrospective processes which monitor the degree to which resultant bodily actions conform with predictive forward models regarding how the brain *expected* those bodily actions to be performed. Evidence points towards a number of psychiatric disorders such

as schizophrenia and psychosis as resulting in part from pathologies within brain mechanisms governing the *how* component of decision-making and the sense of agency. In particular, the functioning sense of agency appears crucial for distinguishing the self from the environment, and for correctly attributing phenomena caused by one's own actions as contrasted with phenomena caused by other people or things – *i.e.,* appreciating the nature and consequences of one's own actions. In turn, it is trite that mental illness – and not least psychotic disorders and instances of psychosis – is grossly overrepresented in prisons as contrasted with the general population. It is submitted that such criminal states of mind as are the subject of *mens rea* – intention, recklessness, knowledge and beliefs, *etc.* – may arise in the minds of afflicted individuals without them appreciating the nature and consequences of their resultant actions. Again, this is something that the investigation into subjective mental states is not necessarily capable of distinguishing with accuracy.

Two elements of the *when* component of decision-making were considered in particular within the thesis. One aspect is the literal decision of when to initiate any given action, in which regard evidence points towards the SMA region of the brain as being responsible for initiating an intentional decision to act into a volitional physical action. Of particular note, the neuroscientific evidence suggests that the brain decides *when* to initiate a previously determined volitional action *prior* to an individual becoming consciously aware of that particular or intention; indeed, it is further suggested that conscious awareness of a decision to act likely results from the associated motor preparation for that act, and not the actual decision itself. The legal relevance of this follows because current conceptions of legal responsibility require the *mens rea* and *actus reus* of an offence to coincide, premised on the presumption that people possess direct conscious control over their decisions and subsequent actions. However, this presumption becomes untenable if the brain decides both *what* to do and *when* to initiate that choice unconsciously, thereby precluding the possibility of conscious control or intervention.

The second element of the *when* component of a decision concerns the more abstracted question of the point in time when a given decision to act is reached, in particular as it relates to conscious deliberation. The thesis considered the seminal research of Benjamin

Libet and numerous corroborative replications and follow-up studies, which together strongly suggest that the brain decides *what* to do *prior* to an individual becoming consciously aware of that decision. Amongst the evidence considered are various modern fMRI studies demonstrating that the decision of *what* to do can be reliably predicted from activity in the brain preceding conscious awareness of any decision by several seconds – which is a notably long time in neuroscience. This further supports findings from the discussion of the *what* component that decisions of what to do are reached unconsciously by the brain and only enter into conscious awareness later, again undermining the proposition within the legal concept of volition that people possess direct conscious control over the outcome of their *decisions* to act. Nevertheless, this discussion left the door open to the possibility that consciousness provides a final veto over such decisions that are made unconsciously.

This door is subsequently closed from the discussion of the *whether* component of decision-making. Here, again, evidence suggests that the brain unconsciously reaches a final decision as to *whether* or not to initiate or veto a previously (unconsciously) decided and planned action into actual bodily movements. Thus, each of *what* to do, *how* to do it, *when* to do it, and *whether* or not to go ahead and do it are first reached unconsciously, with conscious awareness arising second. This is not necessarily surprising upon reflection; the general neuropsychological commitment to causal determinism and the rejection of Cartesian duality means that consciousness itself cannot be an uncaused process – there is no homunculus in the brain which makes decisions independently of activity in the brain itself.

In a chain of conscious deliberation, the decision to do *x* is first reach unconsciously before arising to the level of conscious awareness; next, a decision not to do *x* might be made and replaced with a decision to do *y*, first unconsciously and second arising to the level of conscious awareness; and, again, the next decision not to do *y* might be made and replaced with a decision to do *z*, first unconsciously and second arising to the level of conscious awareness. Whilst representations reaching the level of conscious awareness may, and even likely do, feed back into the unconscious decision-making mechanisms, conscious thought itself does not appear to have *direct control* over the output of those

unconscious decision-making processes. People almost certainly do not directly and consciously control the outcome of decision-making mechanisms in the brain – you cannot consciously *force* those unconscious processes to reach a particular decision to do *x*.

The implications of this for legal responsibility are multiple. With regards to the concept of volition, the presumption that people have *conscious control* over their *decisions* and actions is almost certainly false. That is not to say that people are unable to control their actions, which they clearly can. Rather, the brain's self-control mechanisms are largely unconscious and automatic, as with the significant majority of its functioning. Indeed, there is a rational logic to this order; a process of conscious deliberation *first requires* some manner of control so that people simply do not select the first option that comes to mind when making a decision, but continue deliberating to consider other options also. Thus, the very process of *conscious* decision-making is premised upon some prior degree of self-control; *not* the opposite way around whereby conscious thought is the mechanism which provides that self-control. The more accurate statement, therefore, is that people possess an "ordinary" capacity (*i.e.,* via automatic processes) to control physical bodily actions such that they conform with unconsciously reached decisions to act.

A further critical implication concerns the focus of the law on subjective mental states. Specifically, subjective states such as intention are deemed relevant because the law assumes that an intended action has been *consciously* chosen, and that the individual can consciously control their decisions in order to choose otherwise. Again, this assumption is most likely false, as the overwhelming body of research considered in this thesis together suggests that the overall decision of what to do – *i.e.,* the arising of some criminal intention, dishonesty, recklessness, knowledge of belief, *etc.* – occurs as a result of automatic and unconscious processes. Again, any role that conscious deliberation has in this process must necessarily be secondary to the unconscious processes that given rise to the conscious experience. Thus, the law significantly rests responsibility for action on a factor which itself appears quite arbitrary and outside of an individual's direct conscious control – that is, on the appearance or not of a given subjective state of mind.

The final chapter of Part One of the thesis considered the *why* component of a decision. On the one hand, it is trite that a person's *motive* for committing a given criminal action is *prima facie* irrelevant to the question of responsibility. On the other hand, it is impossible as a matter of practicality for the law – or any courtroom, judge or jury – to peer inside the subjective mind of another. Consequently, the investigation of people's reasons for their decisions becomes a nonetheless inescapable element of the criminal trial, as the reasons offered form the basis of inferring or imputing the requisite subjective state of mind. However, the discussion of the *why* component revealed that subjective recollection of the *genuine* reasons underlying decisions to act is notoriously unreliable. People have generally poor subjective access to such genuine reasons whilst the brain is readily capable of confabulating *post hoc* explanations; moreover, people have a generally poor ability to subjectively distinguish their own confabulations from genuine reasons, and a poorer ability still to objectively distinguish the confabulations of others. Therefore, again, reliance upon proof subjective mental states that can only be demonstrated indirectly appears to be a particularly *un*reliable basis upon which to rest legal responsibility.

The above notwithstanding, it is trite that people are capable of recognising good and bad reasons for action and applying those reasons to decisions to act one way or another. In this regard, intuitionist models of decision-making permit various routes by which rationality and reason is preserved within decisions. For example, it is hypothesised that moral intuitions are learned in the first place through education and experience which teaches rational moral principles – *i.e.,* an intuition becomes the automatic application of a nonetheless rational principle. Meanwhile, reflection, argumentation, persuasion and other manners of continuing moral education update such intuitions, providing reasons into the automatic decision-making processes via through feedback loops. Thus, the automaticity or unconsciousness of decision-making processes does not necessarily preclude them from being responsive to reason; only that such responsiveness occurs beneath the level of consciousness. This point is critical to the legal concept of volition, one half of which consists of the capacity for people to respond to reason in their decision-making. This capacity remains significantly unaltered by the research considered in the present thesis.

A consistently recurring question that arises from the present research asks what role consciousness might play in decision-making processes which, the evidence strongly suggests, are fundamentally automatic and unconscious. Whilst the thesis does not attempt to substantively investigate either of the hard or soft problems of consciousness, two hypotheses are offered for completeness within the discussion on decision-making in particular. First, it is considered that reaching even the simplest decision at least requires resolution of each of the *what*, *how*, *when* and *whether* components (albeit not *necessarily* the *why* component), which appear to be processed in parallel by the brain. Meanwhile, physical action is significantly serial in the sense that people can only meaningfully attend to a limited number of things at one time, and can only properly carry out one action at a time with any great competence. Thus, it is hypothesised that consciousness may provide a necessary interface through which the brain's countless parallel processes (which together extend beyond the five components identified *solely* in relation to decisions to act) are serialised into a unified experience that can be translated into serial action in the world.

Second, despite the apparent automaticity of decision-making processes, conscious deliberation is nonetheless a phenomenon that ostensibly appears to impact upon, alter and change decisions to act, as well as other types of decisions and judgments. In this regard, it is proposed that *conscious* deliberation of any particular subject, topic or action provides both time and mental resources to those automatic decision-making processes. Thus, by deliberating consciously, people do not simply select the first decision option which arises in the mind, but instead allow further time for other options to similarly arise, and for the relevant competing networks of representative neurons to gather more "evidence", thereby arriving a more accurate or considered decision that has taken a greater number of alternative options and associated evidence into account.

Similarly, the process of *conscious* deliberation may allow for greater mental "resources" such as oxygen and blood sugars to be devoted to the relevant brain regions involved in the process of conscious decision-making. Indeed, so much is implied by the use of fMRI which measures changes associated with blood flow in the brain and extrapolates higher or lower levels of activation from these changes. Thus, even if the possibility of direct,

conscious control over automatic brain processes is refuted, consciousness can nonetheless exhibit an indirect causal role in decision-making and is likely not, therefore, epiphenomenal.

*Second Objective*

The second objective of the thesis follows from the conclusions of the first, to appropriately reformulate the current conception of legal responsibility and *mens rea* in particular, taking into account the implications of current scientific research on decision-making, reasoning and volitional control. Part Two of the thesis therefore begins by deconstructing the current conception of *mens rea* according to the neuropsychological implications previously considered. Starting with the concept of volition, the capacity for being responsive to reason remains unaltered as both evidence and theory readily demonstrate the brain's ability to recognise and respond to good or bad reasons for a particular decision or judgment.

The capacity for conscious control over decisions and actions does require some revision, although the overall consequences for the legal concept of volition are not significant. Specifically, the notion of *conscious* control over *decisions* and action ought to be replaced with the notion of ordinary control over *action*. This is meant to capture to idea that the neurotypical brain contains automatic mechanisms which operate to ensure that the body physically performs the actions that are intended and instructed by the brain in order to implement (unconsciously reached) decisions to act. In this regard, most people commonly experience the ability to accurately reach and grasp a glass as intended, rather than always knocking it over, even if the decision to reach for the glass has been reached unconsciously, and even if the specific processes by which such physical control is exerted operate automatically.

Crucially for the purpose of the present thesis, this revision of the capacity of self-control does not significantly interfere with the legal presumption of volition. As volition is presumed in law for all adults, volitional control becomes relevant to such defences as insanity, automatism and loss of control, *etc.,* where requisite self-control is lacking. These defences do not contain any explicit requirement that the relevant self-control is

indeed *conscious*, only that it exists; in which case referring to "ordinary" self-control does not substantively interfere with the operation of the concept of volition (although it does have implications for retributive theories of punishment, discussed below).

Without doubt, the most significant implication of the neuropsychological research concerns reliance upon subjective mental states as a touchstone of legal responsibility. Following from the conclusions under the first objective of the thesis, subjective mental states represent an unreliable and unsafe means of ascribing responsibility. Intentions (and other subjective mental states) can be exogenously triggered and automatically processed to the point of physical action without any necessary opportunity for conscious intervention. They may be arbitrary insofar as they do not necessarily result from deliberative processes or reflect an actual considered opinion, belief or desire of the particular agent. They are unreliably recalled due to poor introspective access to the *operative* reasons behind decisions, and even less reliably proven through the objective observation of third parties. Explanations for decisions and actions are readily confabulated *post hoc*, with neither the individual confabulator nor independent third parties possessing any particularly reliable means of distinguishing said confabulation from genuine explanation.

Instead, the thesis proposes replacing proof of subjective mental states with proof of the capacity to appreciate the nature and consequences of one's actions, as administered through a novel hybrid objective / subjective test. Regarding the objective limb, the previously subjective variations of *mens rea* are replaced with entirely objective definitions which refer to either the relative certainty with which a given criminal action produces a prohibited consequence (intention and recklessness), the relative certainty with which a relevant state of affairs exists (knowledge, belief, suspicion and dishonesty), or the reasonableness of conduct (negligence). Next, the hybrid objective / subject test applies an expected standard of reasonableness to the defendant's particular subjective circumstances and characteristics, asking whether or not it is reasonable to expect any person sharing those subjective elements to appreciate the nature and consequences of their actions as they relate to the objectively defined *mens rea*. The relevant circumstances that may be introduced into this test are those which might impact upon any of the three

capacities identified for responsibility – that is, the capacity of appreciating the nature and consequences of one's actions, and the two volitional capacities of being responsive to reason and exercising ordinary self-control. Thus, it is proposed that the entire concept of *mens rea* is rested upon proof or disproof of these three capacities, rather than proof of the existence of a particular subjective state of mind at the time of the alleged offending.

The capacity-based approach to *mens rea* is defended on a number of grounds. Principally, it is submitted that the three identified capacities are both necessary and sufficient for legal rules to fulfil their essential purpose of guiding behaviour towards or away from desired ends. First, it is proposed that criminal laws in particular exist to identify conduct that has been prohibited by a society and compel and coerce people from engaging in that conduct; second, that a person possessing the three identified capacities has all that is required in theory to conform their behaviour with the law; third, that a person who commits a criminal act whilst in possession of the three capacities must have been caused to so act by factors which overwhelmed the fact of that action's illegality in their decision-making process; and fourth, that the relevant criminal prohibition and its purpose of compelling behaviour would be undermined if those overwhelming factors could not be addressed through the imposition of responsibility and subsequent punishment. To extrapolate further, it is submitted that neither a legal prohibition nor the threat of punishment can reasonably be expected of operate on an individual lacking one or more of the three capacities. Such an individual would be unable to either recognise criminal prohibition as a good reason for factoring into their decision-making; would be unable to control their physical actions to ensure that they comply with a decision to follow the law; or would be unable to appreciate how and why a given decision and action breached that prohibition in the first place.

A number of further advantages of the capacity-based approach to responsibility are claimed. In greatest contrast to requiring proof of subjective mental states, proof of the three relevant capacities can be achieved with far greater certainty and reliability. Unlike a particular state of mind which is momentary and fleeting, mental capacities are considerably more stable, whilst proof of their existence and operation is more readily ascertainable from the third-person perspective of a courtroom, jury and judge. What is

more, deficiencies in any of the claimed capacities commonly result in more lasting behavioural effects which may independently be observed as an indication of the underlying lack of capacity. Whilst people may not always actively engage these capacities in every decision to act, their availability for exercise and application is again more stable and persistent, and less arbitrary or liable to undue interference from extraneous sources than the temporary activation of a mere state of mind. Moreover, the three capacities provide ready targets for the subsequent intervention of the criminal law in cases of both responsibility and non-responsibilty. For example, deterrent punishments target a guilty defendant's responsiveness to reason by providing good reasons to avoid criminal conduct. Meanwhile, rehabilitation aims to identify and remedy causes of criminal conduct and, as such, may appropriately be targeted towards both responsible, guilty defendants and those who have been found not-responsible but still require intervention of the State as necessary for the broader safety and security of society.

The thesis subsequently tested the efficacy of the capacity-based theory of responsibility against leading jurisprudence concerning each form of *mens rea*. In each of these hypothetical applications, it is submitted that the capacity-based approach was readily capable of achieving either the same outcome as the existing law or, better yet, was capable of returning greater justice. For example, in cases where the existing law has significantly struggled to achieve justice through the application of entirely subjective or entirely objective approaches to *mens rea*, it is submitted that the hybrid objective / subjective approach is capable of reaching the fair and just outcome. This is most clearly demonstrated in relation to the *mens rea* of recklessness and dishonesty in particular where, facing such struggles as described, UK law has in fact switched numerous times from endorsing subjective tests to objective tests, or *vice versa*.

In a similar vein, the thesis tested the capacity-based approach against legal defences, again demonstrating how justice is achieved through the hybrid objective / subjective test. In this regard, it is submitted that all legal defences are fundamentally underpinned by the claim that one or more of the defendant's relevant capacities for responsibility was significantly diminished or abrogated entirely in the circumstances. Indeed, it is submitted that this is what identifies any given claim as a valid legal defence in the first place, and

is readily captured within the capacity-based theory here presented. Therefore, again, it is submitted that the hybrid objective / subjective, capacity-based approach to legal responsibility achieves either as just and fair an outcome as the existing law, or it manages to achieve greater justice. Most pertinent to the central research question posed in this thesis, the resulting theory of responsibility is entirely rational in a deterministic universe, principally because it facilitates the very teleology of legal rules rather than being premised on an individual's *metaphysically* free choices.

*Third Objective*

The third objective of the thesis consisted of placing the proposed capacity-based theory of responsibility within its broader philosophical background, and suggesting key legal developments that are implied by the present research. In this regard, the thesis rejects moral blame as the central principle underpinning the ascription of responsibility for action. It is submitted that there is no *necessary* connection between criminal proscriptions and morality; many instances are forthcoming across societies, cultures and histories where the law prohibits conduct that many consider to be perfectly moral or, at least, amoral, whilst a greater number of instances can be found where the criminal law fails to regulate self-evidently immoral conduct. Reliance upon moral blame as the central principle underpinning responsibility requires a strongly relativistic moral stance that a large proportion of society will ultimately find unpalatable. Furthermore, moral blame readily denotes a position of metaphysical freedom whereby people are responsible for their actions because they "could have chosen otherwise"; this is both incompatible with the presumption against metaphysical free will upon which the present thesis is based, and is largely incommensurate with the neuropsychological position denying the online, direct, conscious control over the outcome of decisions to act.

The thesis proposes replacing conceptions of moral blame with the concept of "reasonableness" as the central principle underpinning legal responsibility for criminal acts. Drawing from the seminal work of H. L. A. Hart, it is submitted that criminal prohibitions exist first and foremost to announce that certain actions are not tolerated within a society and, through the backing of punishment, to secure that fewer of those prohibited actions are committed. Thus, legal (and, indeed, moral) rules operate at the

most fundamental level to set the boundaries of reasonable, acceptable behaviour from unreasonable behaviour that a given society will not tolerate. This viewpoint is not plagued by allegations of moral relativism because it is far more readily appreciable how and why different societies, cultures and histories can reach different perspectives on reasonable and unreasonable conduct, without having to abandon belief in or support from moral principles. Ascriptions of responsibility based on unreasonableness can readily made without needing to either controversially defend one moral perspective over another or abandon moral principles altogether to relativism. Moreover, whereas the courts regularly disclaim that they are engaged in a process of *moral* adjudication, the concept of reasonableness is ubiquitous throughout both the criminal and civil law as a standard by which people are judged and against which the courts are eminently more experienced and competent to adjudicate conduct.

One of the key implications of the rejection of moral blame as underpinning responsibility is the subsequent rejection of retributivism as an acceptable theory of punishment, because retributivism is itself fundamentally premised on the existence of moral blame and, in contrast to the present consequentialist theory of responsibility, is entirely retrospective. As with the concept of moral blame itself, retributivism relies on a metaphysical free will that is denied in the presumption of determinism underpinning the present theory, and is largely incommensurate with the body of neuropsychological research reviewed in Part One of the thesis. Nevertheless, it is submitted that the remaining key theories of punishment – incapacitation, deterrence, rehabilitation, restitution and expressivism – are each justifiable on consequentialist grounds that are supported by the present consequentialist theory of responsibility. What is more, retribution argues that punishment is justified in its own right as a moral good and regardless of whatever negative consequences might ensue, applying a commutative proportionality based on *lex talionis* that is more commensurate with the civil law of contract on tort. Conversely, the proposed consequentialist theory of responsibility implies distributive proportionality in punishment which is both theoretically more appropriate to the criminal law, and which obtains stronger moral, practical, and even economic support.

Alongside the rejection of retributivism, the capacity-based theory of punishment explains the fundamental rationale underlying legal defences. This not only renders different defences more rationally consistent when considered together, but it also provides the means to identifying new defences. Specifically, defences may be raised and subsequently formalised when circumstances or individual characteristics are successfully pleaded as significantly diminishing of abrogating entirely one or more of the crucial capacities for responsibility. To this end, the thesis proposed a partial defence of addiction that is novel to UK law and fully subsumed within the consequentialist rationale of the thesis. Thus, addiction is recognised for having often readily identifiable causes which not only justify a partial reduction in individual responsibility but also provide targets for the response of the criminal justice system and the imposition of rehabilitation in particular. Furthermore, addiction can readily be identified as impacting upon the three crucial capacities of defendants and, in particular, the capacities of being responsive to reason and exercising ordinary self-control which underlie volition. That notwithstanding, the proposed defence is modelled on the existing partial defence of diminished responsibility, recognising the principle that addiction generally does not result in a complete abrogation of volition.

The final chapter of the thesis considered the proposed theory of responsibility within a wider moral context, engaging more directly with the philosophical debate between free will, determinism and responsibility. The capacity-based approach to responsibility can readily be generalised to govern responsibility for actions generally and is not necessarily confined to a legal context. Indeed, it is submitted that the principle of responsibility is itself amoral and not strictly legal *per se*; rather, a person can be considered responsible for any decision to act when that decision is made whilst in possession of the three mental capacities, irrespective of whether the resulting action breaches a legal, moral, administrative, or other type of rule. This follows the teleology of rules themselves which exist to identify desired and undesired conduct and guide resulting behaviour accordingly, and the teleology of punishments which exists in order to enforce rules. In this regard, it is argued that the concept of responsibility necessarily attaches to *decisions to act*.

The philosophical discussions concluding the thesis aim to further justify the capacity-based theory of responsibility in the absence of free will. The discussion on persuasion, manipulation, coercion and compulsion demonstrates why causal determinism does not defeat the concepts of legal or moral responsibility; responsibility is not absolved because a decision results from prior causes as, indeed, *all* decisions result from prior causes. Rather, it is submitted, absolution arises because a decision to act has been taken without the crucial capacities of responding to reason, ordinary self-control and appreciating the nature and consequences of one's conduct. In this regard, both acts of persuasion and coercion can amount to a necessary and sufficient cause of a person's subsequent behaviour, but only coercion absolves responsibility because it undermines or overpowers volition. It is not the chain of causation that is relevant to responsibility, therefore, but the availability of crucial capacities for responsible action. This point is further demonstrated by engaging with the famous counterexamples provided by Harry Frankfurt, who argues that the principle of alternate possibilities – (one of the hallmarks of metaphysical free will) – is not a necessary prerequisite for responsibility.

Finally, in further support of the claimed relevance of the crucial capacities, the thesis demonstrates how the proposed theory of responsibility can support either or both of the hard- and soft-line replies to Derk Pereboom's manipulation argument against responsibility in a deterministic universe. These discussions are, again, most pertinent to the central research question posed in this thesis, demonstrating how and why the resulting capacity-based theory provides an entirely rational means of ascribing responsibility for action in a deterministic universe absent of metaphysically free decision-making.

# 15. Bibliography

## ***Published Books and Individual Chapters***


al'Absi M. (ed.), *Stress and Addiction: Biological and Psychological Mechanisms* (Academic Press 2007).

Alexander L., Ferzan K. K. and Morse S., *Crime and Culpability: A Theory of Criminal Law* (Cambridge University Press 2009).

Allen C., Bekoff M. and Lauder G. V., *Nature's Purposes: Analyses of Function and Design in Biology* (Massachusetts Institute of Technology Press 1998).

Allen J. J. and Anderson C. A., 'General aggression model' in Rössler P. and Hoffner C. A. (eds.), *The International Encyclopedia of Media Effects* (John Wiley & Sons 2017).

Allen M., *Criminal Law* (14th ed. Oxford University Press 2017).

Allen M. and Edwards I., *Criminal Law* (15th ed. Oxford University Press 2019).

Allport G. W., *The Nature of Prejudice* (Addison-Wesley 1954).

American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders* (5th ed. American Psychiatric Association 2013).

Anderson C. A. and Carnagey N. L., 'Violent evil and the general aggression model' in Miller A. G. (ed.), *The Social Psychology of Good and Evil* (1st ed. The Guildford Press 2004).

Anderson C. A., Carnagey N. L., Flanagan M., Benjamin A. J., Eubanks J. and Valentine J. C., 'Violent video games: Specific effects of violent content on aggressive thoughts and behavior' in Zanna M. P. (ed.), *Advances in Experimental Social Psychology: Vol. 36* (Elsevier Academic Press 2004).

Anderson C. A. and Huesmann L. R., 'Human aggression: A social-cognitive view' in Hogg M. A. and Cooper J. (eds.), *The SAGE Handbook of Social Psychology* (SAGE Publications 2007).

Apel R. and Nagin D. S., 'General deterrence' in Tonry M. H. (ed.), *The Oxford Handbook of Crime and Criminal Justice* (Oxford University Press 2011).

Aquinas T., *Summa Theologica: Volume I – Part I* (Fathers of the English Dominican Province (trns.), Cosimo Inc. 2007).

Aristotle, *Nicomachean Ethics* (Bartlett R. C. and Collins S. D. (trns.), University of Chicago Press 2011).

Ashworth A., 'Criminal law, human rights and preventative justice' in McSherry B., Norrie A. and Bronitt S. (eds.), *Regulating Deviance: The Redirection of Criminalization and the Futures of Criminal Law* (Hart Publishing 2009).

Ashworth A., *Principles of Criminal Law* (5th ed. Oxford University Press 2006).

Ashworth A. and Horder J., *Principles of Criminal Law* (7th ed. Oxford University Press 2013).

Augustine (of Hippo), *The City of God* (Dods M. (trns.), Hendrickson Publishers 2009).

Badar M. E., *The Concept of Mens Rea in International Criminal Law: The Case for a Unified Approach* (Hart Publishing 2013).

Banks C., *Criminal Justice Ethics: Theory and Practice* (5th ed. SAGE Publications 2020).

Bargh J. A., 'Automatic information processing: Implications for communication and affect' in Donohew L., Sypher H. E. and Higgins E. T. (eds.), *Communication, Social Cognition, and Affect* (Psychology Press 1988).

Bargh J. A. and Chartrand T. L., 'The mind in the middle: A practical guide to priming and automaticity research' in Reis H. T. and Judd C. M. (eds.), *Handbook of Research Methods in Social and Personality Psychology* (2nd ed. Cambridge University Press 2014).

Barnhill A., 'What is manipulation?' in in Coons C. and Weber M. (eds.), *Manipulation: Theory and Practice* (Oxford University Press 2014).

Baumeister R. F., Heatherton T. F. and Tice D. M., *Losing Control: How and Why People Fail at Self-Regulation* (Academic Press 1994).

Bazemore G. and Dooley M., 'Restorative justice and the offender: The challenge of reintegration' in Bazemore D. and Schiff M. (eds.), *Restorative Community Justice: Repairing Harm and Transforming Communities* (Anderson Publishing 2001).

Beatty D. M., *The Ultimate Rule of Law* (Oxford University Press 2004).

Beccaria C., *On Crimes and Punishments and Other Writings* (Bellamy R. (ed.), Cambridge University Press 1995).

Bem D. J., 'Self-perception theory' in Berkowtiz L. (ed.), *Advances in Experimental Social Psychology* (Academic Press 1972).

Bennett C., 'Punishment and rehabilitation' in Ryberg J. and Corlett J. R. (ed.), *Punishment and Ethics: New Perspectives* (Palgrave Macmillan 2010).

Bennett M. R. and Hacker P. M. S., *History of Cognitive Neuroscience* (Wiley-Blackwell 2008).

Bennett T. and Holloway K., *Understanding Drugs, Alcohol and Crime* (Open University Press 2005).

Bentham J., *The Rationale of Punishment* (McHugh J. T. (ed.), *Prometheus Books* 2009).

Billig M., *Arguing and Thinking: A Rhetorical Approach to Social Psychology* (Cambridge University Press 1996).

Bingham T. H., 'The rule of law' (2007) 66(1) *Cambridge Law Journal* 67.

Blackstone W., *Commentaries on the Laws of England in Four Books, Volume 2* (George Sharswood (ed.), J. B. Lippincott Co. 1875).

Blackstone, *Commentaries on the Laws of England in Four Books, Volume 4* (Edward Christian (ed.), A. Strahan and W. Woodfall 1795).

Blair C. and Ursache A., 'A bidirectional model of executive functions and self-regulation' in Vohs K. D. and Baumeister R. F. (eds.), *Handbook of Self-Regulation: Research, Theory, and Applications* (2nd ed. The Guildford Press 2011).

Bogg A. and Herring J., 'Addiction and responsibility' in Herring J., Regan C., Weinberg D. and Withington P. (eds.), *Intoxication and Society: Problematic Pleasures of Drugs and Alcohol* (Palgrave Macmillan 2013).

Bohlander M., 'From *Marx* to *Majewski*: A review of the law on voluntary intoxication in the former German Democratic Republic' in Livings B., Reed A. and Wake N. (eds.), *Mental Condition Defences and the Criminal Justice System: Perspective from Law and Medicine* (Cambridge Scholars Publishing 2015).

Boi L., *The Quantum Vacuum: A Scientific and Philosophical Concept, from Electrodynamics to String Theory and the Geometry of the Microscopic World* (Johns Hopkins University Press 2011).

Bonta J., Jesseman R., Rugge T. and Cormier R., 'Restorative justice and recidivism' in Sullivan D. and Tifft L. (eds.), *Handbook of Restorative Justice: A Global Perspective* (Routledge 2006).

Boswell J., *The Life of Samuel Johnson LL. D.: Including a Journal of a Tour to the Hebrides* (Croker J. W. (ed.), George Dearborn & Co. 1833).

Bracton H., *Bracton on the Laws and Customs of England* (Thorne S. E. (trns.) Belknap Press 1968).

Braithwaite J., 'Restorative justice' in Tonry M. H. (ed.), *The Handbook of Crime and Punishment* (Oxford University Press 1998).

Brink D. O., *Fair Opportunity and Responsibility* (Oxford University Press 2021).

Brooks T., *Deterrence* (Ashgate Publishing 2014).

Brooks T., *Punishment* (Routledge 2012).

Brooks T., *Punishment: A Critical Introduction* (2nd ed. Routledge 2021).

Brown A. L. and DeLoache J. S., 'Skills, plans, and self-regulation' in Siegler R. (ed.), *Children's Thinking: What Develops?* (Lawrence Erlbaum Associates 1978).

Brown D., *Human Universals* (McGraw-Hill Companies 1991).

Brown S., Esbensen F.-A. and Geis G., *Criminology: Explaining Crime and Its Context* (8th ed. Elsevier Inc 2013).

Bruinsma G. J. N., 'Classical theory: The emergence of deterrence theory in the Age of Enlightenment' in Nagin D. S., Cullen F. T. and Jonson C. L. (eds.), *Deterrence, Choice, and Crime: Contemporary Perspectives* (Routledge 2018).

Burns K., Loughnan A., Lunney M. and Willis S., 'Australia: A land of plenty (of legislative regimes)' in Dyson M. (ed.), *Comparing Tort and Crime: Learning from Across and Within Legal Systems* (Cambridge University Press 2015).

Bushman B. J., Huesmann L. R. and Whitaker J. L., 'Violent media effects' in Nabi R. L. and Oliver M. B. (eds.), *The SAGE Handbook of Media Processes and Effects* (SAGE Publications 2009).

Bushway S. and Stoll M. A. (eds.), *Do Prisons Make Us Safer: The Benefits and Costs of the Prison Boom* (Russell Sage Foundation 2009).

Bybee and Suzanne Fleischman (eds.), *Modality in Grammar and Discourse* (John Benjamins Publishing 1995).

Bytsenko A. A., Cognola G., Elizalde E., Moretti V. and Zerbini S., *Analytic Aspects of Quantum Fields* (World Scientific Publishing Co. 2003).

Cantor N. and Kihltrom J. F., *Personality and Social Intelligence* (Prentice-Hall 1987).

Card R. and Molloy J., *Card, Cross & Jones Criminal Law* (22nd ed. Oxford University Press 2016).

Carlen P., 'Crime, inequality and sentencing' in Duff A. and Garland D. (eds.), *A Reader on Punishment* (Oxford University Press 1994).

Carpentier F. R. D., 'Priming' in Rössler P., Hoffner C. A. and van Zoonen L. (eds.), *The International Encyclopedia of Media Effects* (John Wiley & Sons 2017).

Carroll S., *The Big Picture: On the Origins of Life, Meaning and the Universe Itself* (Dutton 2016).

Caruso G. D., ‘Free will skepticism and criminal justice: The public health-quarantine model’ in Nelkin D. K. and Pereboom D. (eds.), *The Oxford Handbook of Moral Responsibility* (Oxford University Press 2022).

Caruso G. D., *Public Health and Safety: The Social Determinants of Health and Criminal Behavior* (ResearcherLinks Books 2017).

Caruso G. D., *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice* (Cambridge University Press 2021).

Chalfin A. J. and Tahamont S., ‘The economics of deterrence: A review of the theory and evidence’ in Nagin D. S., Cullen F. T. and Jonson C. L. (eds.), *Deterrence, Choice, and Crime: Contemporary Perspectives* (Routledge 2018).

Child J. and Ormerod D., *Smith, Hogan, and Ormerod’s Essentials of Criminal Law* (3rd ed. Oxford University Press 2019).

Chomsky N., *Knowledge of Language: Its Nature, Origin, and Use* (Praeger 1986).

Chomsky N., *The Minimalist Program* (Massachusetts Institutes of Technology Press 1995).

Churchland P. S., *Braintrust: What Neuroscience Tells Us about Morality* (Princeton University Press 2011).

Churchland P. S., *Conscience: The Origins of Moral Intuition* (W. W. Norton & Company 2019).

Churchland P. S., ‘Moral decision-making and the brain’ in Illes J. (ed.), *Neuroethics: Defining the Issues in Theory, Practice, and Policy* (Oxford University Press 2006).

Cicero M. T., *The Republic and the Laws* (Rudd N. (trns.), Oxford University Press 1998).

Clore G. L. and Gasper K., ‘Feeling is believing: Some affective influences on belief’ in Frijda N. H., Manstead A. S. R. and Bem S. (eds.), *Emotions and Beliefs: How Feelings Influence Thoughts* (Cambridge University Press 2000).

Coons C. and Weber M., ‘Coercion, manipulation, exploitation’ in Coons C. and Weber M. (eds.), *Manipulation: Theory and Practice* (Oxford University Press 2014).

Copp D., ‘“Ought” implies “can”, blameworthiness, and the principle of alternate possibilities’ in Widerker D. and McKena M. (eds.), *Moral Responsibility and Alternative Possibilities* (Ashgate Publishing 2003).

Cross N., *Criminal Law and Criminal Justice: An Introduction* (SAGE Publications 2010).

Cushman F. and Greene J. D., 'The philosopher in the theater' in Mikulincer M. and Shaver P. R. (eds.), *The Social Psychology of Morality: Exploring the Causes of Good and Evil* (American Psychological Association Press 2011).

Dennett D. C., *Elbow Room: The Varieties of Free Will Worth Wanting* (2nd ed. Massachusetts Institute of Technology Press 2015).

Dessalles J.-L., *Why We Talk: The Evolutionary Origins of Language* (Oxford University Press 2007).

Deutsch M. and Collins M. E., *Interracial Housing: A Psychological Evaluation of a Social Experiment* (University of Minnesota Press 1951).

Devlin P., *The Enforcement of Morals* (Liberty Fund 2010).

Devos T., Huynh Q.-L. and Banaji M. R., 'Implicit self and identity' in Leary M. R. and Tangney J. P. (eds.), *Handbook of Self and Identity* (2nd ed. The Guilford Press 2012).

DeWall C. N., Anderson C. A. and Bushman B. J., 'Aggression' in Tennen H., Suls J. and Weiner I. B. (eds.), *Handbook of Psychology: Vol 5* (2nd ed. John Wiley & Sons 2012).

Dicey A. V., *The Law of the Constitution* (Allison J. W. F. (ed.), Oxford University Press 2013).

Dijksterhuis A. and Bargh J. A., 'The perception-behavior expressway: Automatic effects of social perception on social behavior' in Zanna M. P. (ed.), *Advances in Experimental Social Psychology: Vol. 33* (Academic Press 2001).

Dostoevsky F., *Notes from the Underground* (Garnett C. (trs.), Guignon C. and Aho K. (eds.), Hackett Publishing Company 2009).

Duff R. A., *Answering for Crime: Responsibility and Liability in the Criminal Law* (Hart Publishing 2007).

Duff R. A., *Punishment, Communication, and Community* (Oxford University Press 2003).

Dworkin R., *Taking Rights Seriously* (Bloomsbury Academic 2013).

Eiser J. R., 'A History of Social Judgment Research' in Kruglanski A. W. and Stroebe W. (eds.), *Handbook of the History of Social Psychology* (Psychology Press 2012).

Elvin J. and de Than C., 'The boundaries of the insanity defence: The legal approach where the defendant did not "know that what he was doing was wrong"' in Livings B., Reed A. and Wake N. (eds.), *Mental Condition Defences and the Criminal Justice System: Perspectives from Law and Medicine* (Cambridge Scholars Publishing 2015).

Engle E., 'The general principle of proportionality and Aristotle' in Huppes-Cluysenaer L. and Coelho N. M. M. S. (eds.), *Aristotle and the Philosophy of Law: Theory, Practice and Justice* (Springer Science and Business Media Dordrecht 2013).

Evans J. St. B. T., Newstead S. E. and Byrne R. M. J., *Human Reasoning: The Psychology of Deduction* (Psychology Press 1993).

Evans J. St. B. T. and Over D. E., *Rationality and Reasoning* (Psychology Press 1996).

Faber R. J. and Vohs K. D., 'To buy or not to buy? Self-control and self-regulatory failure in purchase behavior' in Baumeister R. F. and Vohs K. D. (eds.), *Handbook of Self-Regulation: Research, Theory, and Applications* (The Guildford Press 2004).

Faden R. R. and Beauchamp T. L., *A History and Theory of Informed Consent* (Oxford University Press 1986).

Fassin D., *The Will to Punish* (Oxford University Press 2018).

Filbey F. M., *The Neuroscience of Addiction* (Cambridge University Press 2019).

Filbey F. M., Claus E. D. and Hutchinson K. E., 'A neuroimaging approach to the study of craving' in Adinoff B. and Stein E. A. (eds.), *Neuroimaging in Addiction* (John Wiley & Sons 2011).

Fingarette H., *The Meaning of Criminal Insanity* (University of California Press 1972).

Fischer J. M., 'Desert and the Justification of Punishment' in Nadelhoffer T. A. (ed.), *The Future of Punishment* (Oxford University Press 2013).

Fischer J. M. and Ravizza M., *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge University Press 1998).

Fleming G. R. and Scholes G. D., 'Quantum biology: introduction' in Mohseni M., Omar Y., Engel G. and Plenio M. B. (eds.), *Quantum Effects in Biology* (Cambridge University Press 2014).

Fletcher G. P., *Basic Concepts of Criminal Law* (Oxford University Press 1998).

Fletcher G. P., *Rethinking Criminal Law* (Oxford University Press 2000).

Forgas J. P., Baumeister R. F. and Tice D. M., 'The psychology of self-regulation: An introductory review' in Forgas J. P., Baumeister R. F. and Tice D. M. (eds.), *Psychology of Self-Regulation: Cognitive, Affective, and Motivational Processes* (Psychology Press 2009).

Fox D. and Stein A., 'Dualism and doctrine' in Patterson D. and Pardo M. S. (eds.), *Philosophical Foundations of Law and Neuroscience* (Oxford University Press 2016).

Frankfurt H. G., 'Coercion and moral responsibility' in Frankfurt H. G. (ed.), *The Importance of What We Care About: Philosophical Essays* (Cambridge University Press 1998).

Frith C. D., *The Cognitive Neuropsychology of Schizophrenia* (Lawrence Erlbaum Associates 1992).

Fry E. G., *Memoir of the Life of Elizabeth Fry: With Extracts from Her Journal and Letters: Vol. 1* (Fry K. and Creswell R. E. (eds.), Charles Gilpin and John Hatchard & Son 1847).

Funk T. M., *Rethinking Self-Defence: The "Ancient Right's" Rationale Disentangled* (Bloomsbury Publishing 2021).

Gallistel C. R., 'The replacement of general-purpose learning models with adaptively specialized learning modules' in Gazzaniga M. S. (ed.), *The Cognitive Neurosciences* (2nd ed. Massachusetts Institute of Technology Press 2000).

Gazzaniga M. S., *The Mind's Past* (University of California Press 1998).

Gazzaniga M. S., *Who's in Charge? Free Will and the Science of the Brain* (Robinson 2012).

Gazzaniga M. S., Ivry R. B. and Mangun G. R., *Cognitive Neuroscience: The Biology of the Mind* (5th ed. W. W. Norton & Co. 2019).

Gazzaniga M. S. and LeDoux J. E., *The Integrated Mind* (Plenum Press 1978).

Gentile D. A. and Anderson C. A., 'Violent video games: The newest media violence hazard' in Gentile D. A. (ed.), *Advanced in Applied Developmental Psychology. Media Violence and Children: A Complete Guide for Parents and Professionals* (Praeger Publishers 2003).

Gert B., 'Coercion and freedom' in Pennock J. R. and Chapman J. W. (eds.), *Coercion* (Taylor & Francis 1973).

Ginsberg M., *On Justice in Society* (Penguin 1965).

Goeders N. E., 'The hypothalamic-pituitary-adrenal axis and addiction' in al'Absi M. (ed.), *Stress and Addiction: Biological and Psychological Mechanisms* (Academic Press 2007).

Golding M. P., *Philosophy of Law* (Englewood Cliffs 1975).

Goldstein D. S., *Adrenaline and the Inner World: An Introduction to Scientific Integrative Medicine* (Johns Hopkins University Press 2006).

Gottfredson M. R. and Hirschi T., *A General Theory of Crime* (Stanford University Press 1990).

Gottfredson M. R. and Hirschi T., *Modern Control Theory and the Limits of Criminal Justice* (Oxford University Press 2020).

Greene J. D., 'The secret joke of Kant's soul' in Sinnott-Armstrong W. (ed.), *Moral Psychology Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development* (Massachusetts Institute of Technology Pres 2008).

Haggard P., 'Neuroethics of free will' in Illes J. and Sahakian B. J. (eds.), *The Oxford Handbook of Neuroethics* (Oxford University Press 2011).

Haji I., *Moral Appraisability: Puzzles, Proposals, and Perplexities* (Oxford University Press 1998).

Hall J., *General Principles of Criminal Law* (2nd ed. The Lawbook Exchange 2005).

Hameroff S. R., 'Biological feasibility of quantum approaches to consciousness' in van Loocke P. (ed.), *The Physical Nature of Consciousness* (John Benjamins Publishing Company 2001).

Hameroff S. R. and Penrose R., 'Orchestrated reduction of quantum coherence in brin microtubules: A model for consciousness' in Hameroff S. R., Kaszniak A. W. and Scott A. C. (eds.), *Toward a Science of Consciousness: The First Tucson Discussions and Debates – Vol. I* (Massachusetts Institute of Technology Press 1996).

Hamilton D. L., Katz L. B. and Leirer V. O., 'Organizational processes in impression formation' in Hastie R., Ostrom T. M., Ebbesen E. B., Wyer R. S., Hamilton D. L. and Carlston D. E. (eds.), *Person Memory: The Cognitive Basis of Social Perception* (Lawrence Erlbaum Associates 1980a).

Hannigan B., *Company Law* (5th ed. Oxford University Press 2018).

Harris J. A., *Of Liberty and Necessity: The Free Will Debate in Eighteenth-Century British Philosophy* (Oxford University Press 2005).

Hart H. L. A., 'Intention and punishment' in Hart H. L. A. (ed.), *Punishment and Responsibility: Essays in the Philosophy of Law* (2nd ed. Oxford University Press 2008).

Hart H. L. A., 'Legal responsibility and excuses' in Hart H. L. A. (ed.), *Punishment and Responsibility: Essays in the Philosophy of Law* (2nd ed. Oxford University Press 2008).

Hart H. L. A., 'Negligence, *mens rea*, and criminal responsibility' in Hart H. L. A. (ed.), *Punishment and Responsibility: Essays in the Philosophy of Law* (2nd ed. Oxford University Press 2008).

Hart H. L. A., 'Prolegomenon to the principles of punishment' in Hart H. L. A. (ed.), *Punishment and Responsibility: Essays in the Philosophy of Law* (2nd ed. Oxford University Press 2008).

Heckhausen H., *Motivation and Action* (Springer New York 1991).

Herring J., *Criminal Law* (11th ed. Red Globe Press 2019).

Herring J., *Criminal Law: Text, Cases, and Materials* (9th ed. Oxford University Press 2020).

Herring J., *Great Debates in Criminal Law* (4th ed. Red Globe Press 2020).

Herring J., *Medical Law and Ethics* (7th ed. Oxford University Press 2018).

Higgins E. T., 'Knowledge activation: Accessibility, applicability, and salience' in Higgins E. T. and Kruglanski A. W. (eds.), *Social Psychology: Handbook of Basic Principles* (Guilford Publications 1996).

Hirsch A., Bottoms A. E., Burney E. and Wikstrom P. O., *Crime Deterrence and Sentencing Severity* (Hart Publishing 1999).

Hirstein W., *Brain Fiction: Self-deception and the Riddle of Confabulation* (Massachusetts Institute of Technology 2005).

Hodgson D., 'Quantum physics, consciousness, and free will' in Kane R. (ed.), *The Oxford Handbook of Free Will* (Oxford University Press 2011).

Holmes O. W., *The Common Law* (Cosimo Inc 2009).

Horder J., *Ashworth's Principles of Criminal Law* (9th ed. Oxford University Press 2019).

Horst S., *Laws, Mind, and Free Will* (Massachusetts Institute of Technology Press 2011).

Hudson B. A., *Understanding Justice: An Introduction to Ideas, Perspectives and Controversies in Modern Penal Theory* (2nd ed. Open University Press 2003).

Huesmann L. R., Dubow E. F. and Yang G., 'Why it is hard to believe that media violence causes aggression' in Dill K. E. (ed.), *The Oxford Handbook of Media Psychology* (Oxford University Press 2013).

Hyman J., *Action, Knowledge, and Will* (Oxford University Press 2015).

Jackendoff R., *Patterns in the Mind: Language and Human Nature* (Basic Books 1995).

Jaeger L., *The Second Quantum Revolution: From Entanglement to Quantum Computing and Other Super-Technologies* (Springer Nature Switzerland 2018).

James W., *The Principles of Psychology* (MacMillan 1890).

Johansson P., Hall L. and Chater N., 'Preference change through choice' in Dolan R. and Sharot T. (eds.), *Neuroscience of Preference and Choice* (Elsevier Academic Press 2011).

Johnson-Laird P., *How We Reason* (Oxford University Press 2006).

Jones J. W., *The Law and Legal Theory of the Greeks: An Introduction* (Clarendon Press 1956).

Justinian F. P. S., 'The *Lex Aquilia*' in Watson A. (trns.), *The Digest of Justinian: Volume I* (University of Pennsylvania Press 1985).

Kahn S., *Kant, Ought Implies Can, the Principle of Alternative Possibilities, and Happiness* (Lexington Books 2019).

Kahneman D. and Frederick S., 'Representativeness revisited: Attribute substitution in intuitive judgment' in Gilovich T., Griffin D. and Kahneman D. (eds.), *Heuristics and Biases* (Cambridge University Press 2002).

Kahneman D., Slovic P. and Tversky A. (eds.), *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press 1982).

Kane R., 'The contours of contemporary free-will debates (part 2)' in Kane R. (ed.), *The Oxford Handbook of Free Will* (2nd ed. Oxford University Press 2011).

Kant I., *Critique of Pure Reason* (2nd ed., Smith N. K. (trns.) Palgrave Macmillan 2007).

Kant I., *Religion within the Boundaries of Mere Reason: And Other Writings* (Wood A. and Giovanni G. (eds.) Cambridge University Pres 2018).

Kant I., *The Metaphysics of Morals* (Denis L. (ed.), Gregor M. (trns.), Cambridge University Press 2017).

Klein S. B. and Kihlstrom J. F., 'On bridging the gap between social-personality psychology and neuropsychology' in Cacioppo J. T. and Berntson G. G. (eds.), *Foundations in Social Neuroscience* (Massachusettes Institute of Technology 2002).

Koch C., 'Free will, physics, biology, and the brain' in Murphy N., Ellis G. F. R. and O'Connor T. (eds.), *Downward Causation and the Neurobiology of Free Will* (Springer-Verlag Berlin Heidelberg 2009).

Kornhuber H. H., Deecke L., Lang W., Lang M. and Kornhuber A., 'Will, volitional action, attention and cerebral potentials in man: *Bereitschaftspotential* performance-related potentials, direction attention potential, EEG spectrum changes' in W. A. Hershberger (ed.), *Volitional Action: Vol. 62* (North Holland 1989).

Korsgaard C. M., *Creating the Kingdom of Ends* (Cambridge University Press 1996).

Kuhn D., *The Skills of Argument* (Cambridge University Press 1991).

Laan P. H., 'Part-time incapacitation: Probation supervision and electronic monitoring' in Malsch M. and Duker M.(eds), *Incapacitation: Trends and New Perspectives* (Routledge 2016).

Lashley K. S., 'The problem of serial order in behavior' in Jeffress L. A. (ed.), *Cerebral Mechanisms in Behavior* (Wiley 1951).

Lecca P. and Re A., *Theoretical Physics for Biological Systems* (CRC Press 2019).

LeDoux, Donald H. Wilson and Michael S. Gazzaniga, 'Beyond commissurotomy: Clues to consciousness' in Gazzaniga M. S. (ed.), *Handbook of Behavioral Neurobiology Volume 2: Neuropsychology* (Plenum Press 1979).

Leverick F., *Killing in Self-Defence* (Oxford University Press 2006).

Levy N., *Consciousness and Moral Responsibility* (Oxford University Press 2014).

Libet B., Alberts W. W., Wright Jr E. W., Lewis M. and Feinstein B., 'Cortical representation of evoked potentials relative to conscious sensory responses and of somatosensory qualities – in man' in Kornhuber H. H. (ed.), *The Somatosensory System* (Thieme 1975).

Lipsey M. A., Landenberger N. A. and Chapman G. L., 'Rehabilitation: An assessment of theory and research' in Sumner C. (ed.), *The Blackwell Companion to Criminology* (Blackwell Publishing 2004).

Logan G. D., 'On the ability to inhibit thought and action: A users' guide to the stop signal paradigm' in Carr D. D. T. H. (ed.), *Inhibitory Processes in Attention, Memory and Language* (Academic Press 1994).

Loughnan A., *Manifest Madness: Mental Incapacity in the Criminal Law* (Oxford University Press 2012).

Loughnan A. and Wake N., 'Of blurred boundaries and prior fault: Insanity, automatism and intoxication' in Reed A., Bohlander M., Wake N. and Smith E. (eds.), *General Defences in Criminal Law: Domestic and Comparative Perspectives* (Ashgate Publishing 2014).

Loveless J., Allen M., Derry C., *Complete Criminal Law: Text, Cases, and Materials* (7th ed. Oxford University Press 2020).

Lovibond S., *Ethical Formation* (Harvard University Press 2002).

Lunney M., Nolan D. and Oliphant K., *Tort Law: Text and Materials* (6th ed. Oxford University Press 2017).

Mack R. L., *A Layperson's Guide to Criminal Law* (Greenwood Press 1999).

Mackay R. D., 'Diminished responsibility and mentally disordered killers' in Ashworth A. and Mitchell B. (eds.), *Rethinking English Homicide Law* (Oxford University Press 2000).

Mackay R. D., *Mental Condition Defences in the Criminal Law* (Clarendon Press 1995).

Mackenzie M. M., *Plato on Punishment* (University of California Press 1981).

Mackie J. L., 'Retributivism: A test case for ethical objectivity' in Feinberg J. and Coleman J. (eds.)., *Philosophy of Law* (6th ed. Wadsworth 2000).

Malsch M., Alberts W., de Keijser J. and Nijboer H., 'Disqualification from a profession or an office: Nature and actual practice' in Malsch M. and Duker M.(eds), *Incapacitation: Trends and New Perspectives* (Routledge 2016).

Malsch M. and Duker M., 'Introduction' in Malsch M. and Duker M.(eds), *Incapacitation: Trends and New Perspectives* (Routledge 2016).

Marinelli M., 'Dopaminergic reward pathways and effects of stress' in al'Absi M. (ed.), *Stress and Addiction: Biological and Psychological Mechanisms* (Academic Press 2007).

McAuley F. and McCutcheon J. P., *Criminal Liability: A Grammar* (Sweet and Maxwell 2000).

McDowell J. H., *Mind and World* (Harvard University Press 1994).

McFadden J. and Al-Khalili J., *Life on the Edge: The Coming of Age of Quantum Biology* (Crown Publishers 2014).

McNamara T. P., *Semantic Priming: Perspectives from Memory and Word Recognition* (Psychology Press 2005).

Mele A. R., *Autonomous Agents: From Self-Control to Autonomy* (Oxford University Press 1995).

Mele A. R., *Free Will and Luck* (Oxford University Press 2006).

Meltzoff A. N. and Moore M. K., 'Infants' understanding of people and things: From body imitation to folk psychology' in Bermúdez J. L., Marcel A. J. and Eilan N. (eds.), *The Body and the Self* (Massachusetts Institute of Technology Press 1995).

Miceli, 'Deterrence and incapacitation models of criminal punishment: Can the twain meet?' in Harel A. and Hylton K N. (eds.), *Research Handbook on the Economics of Criminal Law* (Edward Elgar Publishing 2012).

Mikhail J. M., *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment* (Cambridge University Press 2011).

Mikhail J. M., 'Moral grammar and intuitive jurisprudence: A formal model of unconscious moral and legal knowledge' in Bartels D. M., Bauman C. W., Skitka L. J. and Medin D. L. (eds.), *Psychology of Learning and Motivation: Moral Judgment and Decision Making* (Academic Press 2009).

Mill J. S., *'On Liberty' and Other Writings* (Collini S. (ed.) Cambridge University Press 1989).

Mischel W., 'Theory and research on the antecedents of self-imposed delay of reward' in Maher B. A. (ed.), *Progress in Experimental Personality Research: Vol 3* (Academic Press 1966).

Mischel W., *The Marshmallow Test: Understanding Self-Control and How To Master It* (Transworld Publishers 2014).

Mitchell, 'Years of provocation, followed by a loss of control' in Zedner L. and Roberts J. V. (eds.), *Principles and Values in Criminal Law and Criminal Justice: Essays in Honour of Andrew Ashworth* (Oxford University Press 2012).

Monaghan N., *Criminal Law Directions* (6th ed. Oxford University Press 2020).

Moore M. S., *Act and Crime: The Philosophy of Action and Its Implications for Criminal Law* (Oxford University Press 2010).

Moore M. S., 'Liberty's constraints on what should be made criminal' in Duff R. A., Farmer L., Marshall S. E., Renzo M. and Tadros V. (eds.), *Criminalization: The Political Morality of the Criminal Law* (Oxford University Press 2014).

Moore M. S., *Placing Blame: A Theory of the Criminal Law* (Oxford University Press 2010).

Morf C. C. and Mischel W., 'The self as a psycho-social dynamic processing system: Toward a converging science of selfhood' in Leary M. R. and Tangney J. P. (eds.), *Handbook of Self and Identity* (2nd ed. The Guilford Press 2012).

Morse S. J., 'Criminal law and addiction' in Pickard H. and Ahmed S. H. (eds.), *The Routledge Handbook of Philosophy and Science of Addiction* (Routledge 2019).

Morse S. J., 'Moral and legal responsibility and the new neuroscience' in Iles J. (ed.), *Neuroethics: Defining the Issues in Theory, Practice, and Policy* (Oxford University Press 2006).

Moskowitz G. B., Li P. and Kirk E. R., 'The implicit volition model: On the preconscious regulation of temporarily adopted goals' in Zanna M. P. (ed.), *Advances in Experimental Social Psychology, Vol. 36* (Elsevier Academic Press 2004).

Murphy J. G., *Kant: The Philosophy of Right* (Mercer University Press 1994).

Narvaez D., 'The social intuitionist model: Some counter-intuitions' in Sinnott-Armstrong W. (ed.), *Moral Psychology Volume 2: The Cognitive Science of Morality: Intuition and Diversity* (Massachusetts Institute of Technology Press 2008).

Neely J. H., 'Semantic priming effects in visual word recognition: A selective review of current findings and theories' in Besner D. and Humphreys G. W. (eds.), *Basic Processes in Reading: Visual Word Recognition* (Lawrence Erlbaum Associates 1991).

Neumann O. and Klotz W., 'Motor responses to nonreportable, masked stimuli: Where is the limit of direct parameter specification?' in Umiltà C. and Moscovitch M. (eds.), *Attention and Performance 15: Conscious and Nonconscious Information Processing* (Massachusetts Institute of Technology Press 1994).

Nold R., Massingale K. and Hodwitz O., 'Justice in ancient Greece and Rome' in Hodwitz O. (ed.), *The Origins of Criminological Theory* (Routledge 2022).

Norrie A., *Crime, Reason and History* (2nd ed. Butterworths 2001).

Norrie A., *Crime, Reason and History: A Critical Introduction to Criminal Law* (3rd ed. Cambridge University Press 2014).

Norrie A., *Law and the Beautiful Soul* (GlassHouse Press 2005).

Nozick R., 'Coercion' in Morgenbesser W. (ed.), *Philosophy, Science and Method: Essays in Honor of Ernest Nagel* (St. Martin's Press 1969).

Ormerod D. and Laird K., '*Ivey v Genting Casinos* – Much ado about nothing?' in Clarry D. (ed.), *The UK Supreme Court Yearbook: Volume 9* (Appellate Press 2019).

Ormerod D. and Laird K., *Smith, Hogan, and Ormerod's Criminal Law* (15th ed. Oxford University Press 2018).

Ormerod D. and Laird K., *Smith, Hogan, and Ormerod's Text, Cases, and Materials on Criminal Law* (13th ed. Oxford University Press 2020).

Ormerod D., Smith J. C. and Hogan B., *Smith and Hogan's Criminal Law* (13th ed. Oxford University Press 2011).

Packer H., *The Limits of the Criminal Sanction* (Stanford University Press 1968).

Padfield N., *Criminal Law* (10th ed. Oxford University Press 2016).

Pereboom D., *Free Will, Agency, and Meaning in Life* (Oxford University Press 2014).

Pereboom D., 'Free will skepticism and criminal punishment' in Nadelhoffer T. A. (ed.), *The Future of Punishment* (Oxford University Press 2013).

Pereboom D., *Living Without Free Will* (Cambridge University Press 2001).

Pereboom D. and Caruso G. D., 'Hard-incompatibilist existentialism: Neuroscience, punishment, and meaning in life' in Caruso G. D. and Flanagan O. J. (eds.), *Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience* (Oxford University Press 2018).

Perkins D. N., Faraday M. and Bushey B., 'Everyday reasoning and the rots of intelligence' in Voss J. F., Perkins D. N and Segal J. W. (eds.), *Informal Reasoning and Education* (Lawrence Erlbaum Associates 1991).

Peters E. M., 'Prison before the prison: The ancient world and medieval worlds' in Morris N. and Rothman D. J. (eds.), *The Oxford History of the Prison: The Practice of Punishment in Western Society* (Oxford University Press 1995).

Petty R. E. and Wegener D. T., 'Attitude change: Multiple roles for persuasion variables' in Gilbert D. T., Fiske S. T. and Lindzey G. (eds.), *The Handbook of Social Psychology Vol. 1* (4th ed. McGraw Hill 1998).

Petty R. E. Wheeler S. C. and Tormala Z. L., 'Persuasion and attitude change' in Millon T. and Lerner M. J. (eds.), *Handbook of Psychology: Volume 5 – Personality and Social Psychology* (John Wiley & Sons 2003).

Plucknett T. F. T., *A Concise History of the Common Law* (5th ed. The Lawbook Exchange 2001).

Pollock F. and Maitland F. W., *The History of English Law Before the Time of Edward I* (2nd ed. The Lawbook Exchange 2008).

Pratt T. C., Cullen F. T., Blevins K. R., Daigle L. E. and Madensen T. D., 'The empirical status of deterrence theory: A meta-analysis' in Cullen F. T., Wright J. P. and Blevins K. R. (eds.), *Taking Stock: The Status of Criminological Theory: Volume 15* (Transaction Publishers 2006).

Rand A., *Philosophy: Who Needs It* (Signet 1982).

Rapp C., 'Aristotle on the moral psychology of persuasion' in Shields C. (ed.), *The Oxford Handbook of Aristotle* (Oxford University Press 2012).

Rapp C., 'Free will, choice, and responsibility (Book III.1-5 [1-7])' in Höffe O. (ed.), *Aristotle's "Nicomachean Ethics"* (Brill 2010).

Raz J., *From Normativity to Responsibility* (Oxford University Press 2011).

Richards G., *Psychology: The Key Concepts* (Routledge 2009).

Rix K. J. B., 'Prizing open the door to justice: Reform of the "wrongfulness limb" of the *M'Naghten* Rules' in Livings B., Reed A. and Wake N. (eds.), *Mental Condition Defences*

*and the Criminal Justice System: Perspectives from Law and Medicine* (Cambridge Scholars Publishing 2015).

Robinson G. and Crow I. D., *Offender Rehabilitation: Theory, Research and Practice* (SAGE Publications 2009).

Rosenberg M., *Conceiving the Self* (Basic Books 1979).

Roskos-Ewoldsen D. R., Klinger M. R. and Roskos-Ewoldsen B., 'Media priming: A meta-analysis' in Preiss R. W., Gayle B. M., Burrell N., Allen M. and Bryant J. (eds.), *Mass Media Effects Research: Advances Through Meta-Analysis* (Routledge 2007).

Rotman E., 'Beyond punishment' in Duff A. and Garland D. (eds.), *A Reader on Punishment* (Oxford University Press 1994).

Rumbold J., *Automatism as a Defence in Criminal Law* (Routledge 2018).

Saner H., *Kant's Political Thought* (Ashton E. B. (trns.) University of Chicago Press 1973).

Sangero B., *Self-Defence in Criminal Law* (Hart Publishing 2006).

Satinover J., *The Quantum Brain: The Search for Freedom and the Next Generation of Man* (John Wiley & Sons 2001).

Sauer H., *Moral Judgments as Educated Intuitions* (Massachusetts Institute of Technology Press 2017).

Schacter D. L., 'Priming and multiple memory systems: Perceptual mechanisms of implicit memory' in Schacter D. L. and Tulving E. (eds.), *Memory Systems* (Massachusetts Institute of Technology Press 1994).

Shakespeare W., *King Lear* (Bate J. and Rasmussen E. (eds.), Macmillan Publishers 2009).

Shakespeare W., *The Tempest* (Raffel B. (ed.), Yale University Press 2006).

Sherry J. L., 'Violent video games and aggression: Why can't we find effects?' in Preiss R. W., Gayle B. M., Burrell N., Allen M. and Bryant J. (eds.), *Mass Media Effects Research: Advances Through Meta-Analysis* (Routledge 2007).

Shute S., 'Knowledge and belief in the criminal law' in Shute S. and Simester A. (eds.), *Criminal Law Theory: Doctrines of the General Part* (Oxford University Press 2002).

Sidharth B. G., *The Universe of Fluctuations: The Architecture of Spacetime and the Universe* (Springer 2005).

Simester A. P., 'A disintegrated theory of culpability' in Baker D. J. (ed.), *The Sanctity of Life and the Criminal Law: The Legacy of Glanville Williams* (Cambridge University Press 2013).

Simester A. P., 'On justifications and excuses' in Zedner L. and Roberts J. (eds.), *Principles and Values in Criminal Law and Criminal Justice: Essays in Honour of Andrew Ashworth* (Oxford University Press 2012).

Simester A. P., Spencer J. R., Stark F., Sullivan G. R. and Virgo G. J., *Simester and Sullivan's Criminal Law: Theory and Doctrine* (7th ed. Hart Publishing 2019).

Skewes M. C. and Gonzalez V. M., 'The biopsychosocial model of addiction' in Miller P. M., Kavanagh D. J., Kampman K. M., Bates M. E., Larimer M. E., Petry N. M., DeWitte P. and Ball S. A. (eds.), *Principles of Addiction: Comprehensive Addictive Behaviors and Disorders: Volume 1* (Academic Press 2013).

Smeesters D., Wheeler S. C. and Kay A. C., 'Indirect prime-to-behavior effects: The role of perceptions of the self, others, and situations in connected primed constructs to social behavior' in Zanna M. P. (ed.), *Advances in Experimental Social Psychology: Volume 42* (Elsevier Academic Press 2010).

Smilansky S., 'Free will, fundamental dualism, and the centrality of illusion' in Kane R. (ed.), *The Oxford Handbook of Free Will* (2nd ed. Oxford University Press 2011).

Smith J. C., *Justification and Excuse in the Criminal Law* (Stevens & Sons Ltd. 1989).

Smith K. J. M., *Lawyers, Legislators and Theorists: Developments in English Jurisprudence 1800 – 1957* (Clarendon Press 1998).

Sompolinsky H., 'A scientific perspective on human choice' in Berger Y. and Shatz D. (eds.), *Judaism, Science, and Moral Responsibility* (Rowman & Littlefield Publishers 2006).

Spain E., *The Role of Emotions in Criminal Law Defences: Duress, Necessity and Lesser Evils* (Cambridge University Press 2011).

Sperber D., 'Metarepresentations in an evolutionary perspective' in Sperber D. (ed.), *Metarepresentations: A multidisciplinary perspective* (Oxford University Press 2000).

Spinoza B., *Ethics: Demonstrated in Geometric Order* (Kisner M. J. (ed.), Cambridge University Press 2018).

Steele J., *Tort Law: Text, Cases, and Materials* (4th ed. Oxford University Press 2017).

Stephen J. F., *A Digest of the Criminal Law* (8th ed. Sweet and Maxwell 1947).

Stephen J. F., *A General View of the Criminal Law of England* (2nd ed. Macmillan 1890).

Stephen J. F., *A History of the Criminal Law of England – Vol II* (Macmillan 1883).

Stephen J. F., *A History of the Criminal Law of England – Vol III* (Macmillan 1883).

Sterelny K., *The Evolved Apprentice: How Evolution Made Humans Unique* (Massachusetts Institute of Technology Press 2012).

Storey T., *Unlocking Criminal Law* (7th ed. Routledge 2020).

Strawson P. F., 'Freedom and resentment' in Stawson P. F. (ed.), *Freedom and Resentment and Other Essays* (Routledge 2008).

Stroud S. R., *Kant and the Promise of Rhetoric* (Pennsylvania State University Press 2014).

Styles E. A., *Attention, Perception and Memory: An Integrated Introduction* (Psychology Press 2005).

Sullivan G. R., 'Knowledge, belief, and culpability' in Shute S. and Simester A. (eds.), *Criminal Law Theory: Doctrines of the General Part* (Oxford University Press 2002).

Tadros V., *Criminal Responsibility* (Oxford University Press 2007).

Tamanaha B. Z., *On the Rule of Law* (Cambridge University Press 2004).

Tamblyn N., *The Law of Duress and Necessity: Crime, Tort, Contract* (Routledge 2017).

Than C. and Elvin J., 'Mistaken private defence: The case for reform' in Reed A. and Bohlander M. (eds.), *General Defences in Criminal Law: Domestic and Comparative Perspectives* (Ashgate Publishing 2014).

Tonry M. H., *Crime and Justice: A Review of Research, Volume 37* (University of Chicago Press 2008).

Tonry M. H., *Thinking about Crime: Sense and Sensibility in American Penal Culture* (Oxford University Press 2004).

Tooby J. and Cosmides L., 'Conceptual foundations of evolutionary psychology' in Buss D. M. (ed.), *Handbook of Evolutionary Psychology* (John Wiley & Sons 2005).

Tooby J. and Cosmides L., 'The theoretical foundations of evolutionary psychology' in Buss D. M. (ed.), *Handbook of Evolutionary Psychology, Volume 1: Foundation* (John Wiley & Sons 2016).

Townsend J. T. and Busemeyer J., 'Dynamic representation of decision-making' in Port R. F. and van Gelder T. (eds.), *Mind as Motion: Explorations in the Dynamics of Cognition* (Massachusetts Institute of Technology 1995).

Turner J. W. C., *Kenny's Outlines of Criminal Law* (16th ed. Cambridge University Press 1952).

Turner J. W. C., *Kenny's Outlines of Criminal Law* (19th ed. Cambridge University Press 1966).

Turner M. A. and Moran N. F., 'Automatism: The ictus, the character, and the law' in Livings B., Reed A. and Wake N. (eds.), *Mental Condition Defences and the Criminal Justice System: Perspectives from Law and Medicine* (Cambridge Scholars Publishing 2015).

Uniacke S., *Permissible Killing: The Self-Defence Justification of Homicide* (Cambridge University Press 1994).

van Inwagen P., *An Essay on Free Will* (Clarendon Press 1983).

Vogt B. A., 'Regions and subregions of the cingulate cortex' in Vogt B. A. (ed.), *Cingulate Neurobiology and Disease* (Oxford University Press 2009).

von Hirsch A., *Censure and Sanctions* (Clarendon Press 1993).

von Hirsch A., Bottoms A. E., Burney E. and Wikström P.-O., *Criminal Deterrence and Sentencing Severity: An Analysis of Recent Research* (Hart Publishing 1999).

Walgrave L., *Restorative Justice, Self-Interest and Responsible Citizenship* (Routledge 2012).

Wallace R. J., *Responsibility and the Moral Sentiments* (Harvard University Press 1994).

Waller B., *Against Moral Responsibility* (Massachusetts Institute of Technology Press 2011).

Webster C. M. and Doob A. N., 'Searching for sasquatch: Deterrence of crime through sentence severity' in Petersilia J. and Reitz K. R. (eds.), *The Oxford Handbook of Sentencing and Corrections* (Oxford University Press 2012).

Wegner D. M., *The Illusion of Conscious Will* (Massachusetts Institute of Technology Press 2018).

Weigend T., 'Sentencing and punishment in Germany' in Tonry M. and Frase R. S. (eds.), *Sentencing and Sanctions in Western Countries* (Oxford University Press 2001).

Wheeler S. C., DeMarree K. G. and Petty R. E., 'The roles of the self in priming-to-behavior effects' in Tesser A., Wood J. V and Stapel D. A. (eds.), *On Building, Defending, and Regulating the Self: A Psychological Perspective* (Psychology Press 2005).

Wicklund R. A. and Gollwitzer P. M., *Symbolic Self-Completion* (Routledge 1982).

Williams G., *Criminal Law: The General Part* (Stevens & Sons Ltd. 1953).

Williams G., *Criminal Law: The General Part* (2nd ed. Stevens & Sons Ltd. 1961).

Wilner D. M., Walkley R. P. and Cook S. W., *Human Relations in Interracial Housing: A Study of the Contact Hypothesis* (University of Minnesota Press 1955).

Wilson W., 'How criminal defences work' in Reed A. and Bohlander M. (eds.), *General Defences in Criminal Law: Domestic and Comparative Perspectives* (Routledge 2016).

World Health Organization, *International Classification of Diseases* (11th ed. World Health Organisation 2019).

Yannoulidis S., *Mental State Defences in Criminal Law* (Routledge 2016).

Yeung N., 'Conflict monitoring and cognitive control' in Ochsner K. N. and Kosslyn S. (eds.), *The Oxford Handbook of Cognitive Neuroscience: Volume 2: The Cutting Edges* (Oxford University Press 2013).

Zaibert L., *Punishment and Retribution* (Ashgate Publishing 2006).

Zee D. S. and Shaikh A. G., 'The neurology of eye movements: From control systems to genetics to ion channels to targeted pharmacotherapy' in Werner J. S. and Chalupa L. M. (eds.), *The New Visual Neurosciences* (Massachusetts Institute of Technology 2014).

## ***Peer Reviewed Journal Articles***

Aarts H., Custers R. and Veltkamp M., 'Goal priming and the affective-motivational route to nonconscious goal pursuit' (2008) 26(5) *Social Cognition* 555.

Aarts H. and Dijksterhuis A., 'Habits as knowledge structures: Automaticity in goal-directed behavior' (2000a) 78(1) *Journal of Personality and Social Psychology* 53.

Aarts H. and Dijksterhuis A., 'The automatic activation of goal-directed behaviour: The case of travel habit' (2000b) 20(1) *Journal of Environmental Psychology* 75.

Aarts H., Gollwitzer P. M. and Hassin R. R., 'Goal contagion: Perceiving is for pursuing' (2004) 87(1) *Journal of Personality and Social Psychology* 23.

Adachi P. J. C. and Willoughby T., 'The effect of violent video games on aggression: Is it more than just the violence?' (2011) 16(1) *Aggression and Violent Behavior* 55.

Adriaanse M. A., Weijers J., de Ridder D. T. D., Huberts J. W. and Evers C., 'Confabulating reasons for behaving bad: The psychological consequences of unconsciously activated behaviour that violates one's standards' (2014) 44(3) *Journal of Social Psychology* 255.

Ahmed S. F., Tang S., Waters N. E. and Davis-Kean P., 'Executive function and academic achievement: Longitudinal relations from early childhood to adolescence' (2019) 111(3) *Journal of Educational Psychology* 446.

Alexander L. and Link B., 'The impact of contact on stigmatizing attitudes toward people with mental illness' (2003) 12(3) *Journal of Mental Health* 271.

Allen M. D., Bigler E. D., Larson J., Goodrich-Hunsaker N. J. and Hopkins R. O., 'Functional neuroimaging evidence for high cognitive effort on the Word Memory Test in the absence of external incentives' (2007) 21(13-14) *Brain Injury* 1425.

Amirthalingam K., 'Caldwell recklessness is dead, long live *mens rea*'s fecklessness' (2004) 67(3) *Modern Law Review* 491.

Ammon K. and Gandevia S. C., 'Transcranial magnetic stimulation can influence the selection of motor programmes' (1990) 53(8) *Journal of Neurology, Neurosurgery, and Psychiatry* 705.

Amodio D. M., Devine P. G. and Harmon-Jones E., 'Individual differences in the regulation of intergroup bias: The role of conflict monitoring and neural signals for control' (2008) 94(1) *Journal of Personality and Social Psychology* 60.

Anderson C. A., 'An update on the effects of playing violent video games' (2004) 27(1) *Journal of Adolescence* 113.

Anderson C. A., 'Effects of violent movies and trait hostility on hostile feelings and aggressive thoughts' (1997) 23(3) *Aggressive Behavior* 161.

Anderson C. A., Benjamin A. J. and Bartholow B. D., 'Does the gun pull the trigger? Automatic priming effects of weapon pictures and weapon names' (1998) 9(4) *Psychological Science* 308.

Anderson C. A. and Bushman B. J., 'Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior: A meta-analytic review of the scientific literature' (2001) 12(5) *Psychological Science* 353.

Anderson C. A. and Carnagey N. L., 'Causal effects of violent sports video games on aggression? Is it competitiveness of violent content?' (2009) 45(4) *Journal of Experimental Social Psychology* 731.

Anderson C. A. and Dill K. E., 'Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life' (2000) 78(4) *Journal of Personality and Social Psychology* 772.

Anderson C. A., Shibuya A., Ihori N., Swing E. L., Bushman B. J., Sakamoto A., Rothstein H. R. and Saleem M., 'Violent video games effects on aggression, empathy,

and prosocial behavior in Eastern and Western countries: A meta-analytic review' (2010) 136(2) *Psychological Bulletin* 151.

Anderson L., Schleien S. J., McAvoy L., Lais G. and Seligmann D., 'Creating positive change through an integrated outdoor adventure program' (1997) 31(4) *Therapeutic Recreation Journal* 214.

Anderson R. A. and Cui H., 'Intention, action planning, and decision making in parietal-frontal circuits' (2009) 63(5) *Neuron* 568.

Aquino K. and Reed A., 'The self-importance of moral identity' (2002) 83(6) *Journal of Personality and Social Psychology* 1423.

Arenson K. J., 'The pitfalls in the law of attempt: A new perspective' (2005) 69(2) *Journal of Criminal Law* 146.

Ariani G., Wurm M. F. and Lingnau A., 'Decoding internally and externally driven movement plans' (2015) 35(42) *Journal of Neuroscience* 14160.

Arnold D. G., 'Coercion and moral responsibility' (2001) 38(1) *American Philosophical Quarterly* 53.

Aron A. R., Robbins T. W. and Poldrack R. A., 'Inhibition and the right inferior frontal cortex: One decade on' (2014) 18(4) *Trends in Cognitive Sciences* 177.

Ashworth A., 'Case Comment: *Andronicou and Constantinou v Cyprus*' (1998) *Criminal Law Review* 823.

Augustinova M., 'Falsification cueing in collective reasoning: Example of the Wason selection task' (2008) 38(5) *European Journal of Social Psychology* 770.

Austin J. L., 'A plea for excuses' (1956-57) 57 *Proceedings of the Aristotelian Society* 1.

Averbeck B. B., Sohn J.-W. and Lee D., 'Activity in prefrontal cortex during dynamic selection of action sequences' (2006) 9(2) *Nature Neuroscience* 276.

Aveyard M. E., 'A call to honesty: Extending religious priming of moral behavior to Middle Eastern Muslims' (2014) 9(7) *PLoS ONE* e99447.

Ayduk O., Mendoza-Denton R., Mischel W., Downey G., Peake P. K. and Rodriguez M., 'Regulating the interpersonal self: Strategic self-regulation for coping with rejection sensitivity' (2000) 79(5) *Journal of Personality and Social Psychology* 776.

Baan F. C., Montanari L., Royuela L. and Lemmens P. H. H. M., 'Prevalence of illicit drug use before imprisonment in Europe: Results from a comprehensive literature review' (2021) 29(1) *Drugs: Education, Prevention and Policy* 1.

Baars B. J., 'The conscious access hypothesis: Origins and recent evidence' (2002) 6(1) *Trends in Cognitive Sciences* 47.

Badre D. and D'Esposito M., 'Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex' (2007) 19(12) *Journal of Cognitive Neuroscience* 2082.

Badre D., Hoffman J., Cooney J. W. and D'Esposito M., 'Hierarchical cognitive control deficits following damage to the human frontal lobe' (2009) 12(4) *Nature Neuroscience* 515.

Ball T., Schreiber A., Feige B., Wagner M., Lücking C. H. and Kristeva-Feige R., 'The role of higher-order motor areas in voluntary movement as revealed by high-resolution EEG and fMRI' (1999) 10(6) *NeuroImage* 682.

Ballard M. E. and West J. R., 'Mortal Kombat™: The effects of violent videogame play on males' hostility and cardiovascular responding' (1996) 26(8) *Journal of Applied Social Psychology* 717.

Balleine B. W., Killcross A. S. and Dickinson A., 'The effect of lesions of the basolateral amygdala on instrumental conditioning' (2003) 23(2) *Journal of Neuroscience* 666.

Bar-Anan Y., Wilson T. D. and Hassin R. R., 'Inaccurate self-knowledge formation as a result of automatic behavior' (2010) 46(6) *Journal of Experimental Social Psychology* 884.

Barden J. and Tormala Z. L., 'Elaboration and attitude strength: The new meta-cognitive perspective' (2014) 8(1) *Social and Personality Psychology Compass* 17.

Bargh J. A., 'The historical origins of priming as the preparation of behavioral responses: Unconscious carryover and contextual influences of real-world importance' (2014) 32(Supp) *Social Cognition* 209.

Bargh J. A., Bond R. N., Lombardi W. J. and Tota M. E., 'The additive nature of chronic and temporary sources of construct accessibility' (1986) 50(5) *Journal of Personality and Social Psychology* 869.

Bargh J. A. and Chartrand T. L., 'The unbearable automaticity of being' (1999) 54(7) *American Psychologist* 462.

Bargh J. A., Chen M. and Burrows L., 'Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action' (1996) 71(2) *Journal of Personality and Social Psychology* 230.

Bargh J. A. and Ferguson M. J., 'Beyond behaviorism: On the automaticity of higher mental processes' (2000) 126(6) *Psychological Bulletin* 925.

Bargh J. A., Lee-Chai A., Barndollar K., Gollwitzer P. M. and Trötschel R., 'The automated will: Nonconscious activation and pursuit of behavioral goals' (2001) 81(6) *Journal of Personality and Social Psychology* 1014.

Bargh J. A., Lombardi W. J. and Higgins E. T., 'Automaticity of chronically accessible constructs in person x situation effects on person perception: It's just a matter of time' (1988) 55(4) *Journal of Personality and Social Psychology* 599.

Bargh J. A., Schwader K. L., Hailey S. E., Dyer R. L. and Boothby E. J., 'Automaticity in social-cognitive processes' (2012) 16(12) *Trends in Cognitive Sciences* 593.

Bargh J. A. and Thein R. D., 'Individual construct accessibility, person memory, and the recall-judgment link: The case of information overload' (1985) 49(5) *Journal of Personality and Social Psychology* 1129.

Barlas Z. and Obhi S. S., 'Freedom, choice, and the sense of agency' (2013) 7 *Frontiers in Human Neuroscience* 514.

Barnes C. M., Lucianetti L., Bhave D. P. and Christian M. S., '"You wouldn't like me when I'm sleepy": Leaders' sleep, daily abusive supervision, and work unit engagement' (2015) 58(5) *Academy of Management Journal* 1419.

Baron M., 'Justifications and excuses' (2005) 2(2) *Ohio State Journal of Criminal Law* 387.

Baron M., 'Manipulativeness' (2003) 77(2) *Proceedings and Addresses of the American Philosophical Association* 37.

Barrett L. F., Tugade M. M. and Engle R. W., 'Individual differences in working memory capacity and dual-process theories of the mind' (2004) 130(4) *Psychological Bulletin* 553.

Bartholow B. D., Anderson C. A., Carnagey N. L. and Benjamin A. J., 'Interactive effects of life experience and situational cues on aggression: The weapons priming effect in hunters and non-hunters' (2005) 41(1) *Journal of Experimental Social Psychology* 48.

Bartholow B. D. and Heinz A., 'Alcohol and aggression without consumption: Alcohol cues, aggressive thoughts, and hostile perception bias' (2006) 17(1) *Psychological Science* 30.

Battaglia-Mayer A., Babicola L. and Satta E., 'Parieto-frontal gradients and domains underlying eye and hand operations in the action space' (2016) 334 *Neuroscience* 76.

Baumann M. A., Fluet M.-C. and Scherberger H., 'Context-specific grasp movement representation in the macaque anterior intraparietal area' (2009) 29(20) *Journal of Neuroscience* 6434.

Baumeister R. F., Bratslavsky E., Muraven M. and Tice D. M., 'Ego depletion: Is the active self a limited resource?' (1998) 74(5) *Journal of Personality and Social Psychology* 1252.

Baumeister R. F., Gailliot M. T., DeWall C. N. and Oaten M., 'Self-regulation and personality: How interventions increase regulatory success, and how depletion moderates the effects of traits on behavior' (2006) 74(6) *Journal of Personality* 1773.

Baumeister R. F., Masciampo E. J. and Vohs K. D., 'Do conscious thoughts cause behavoir?' (2011) 62(1) *Annual Review of Psychology* 331.

Baumeister R. F., Vohs K. D. and Tice D. M., 'The strength model of self-control' (2007) 16(6) *Current Directions in Psychological Science* 351.

Bebbington P., Jakobowitz S., McKenzie N., Killaspy H., Iveson R., Duffield G. and Kerr M., 'Assessing needs for psychiatric treatment in prisoners: 1. Prevalence of disorder' (2017) 52(2) *Social Psychiatry and Psychiatric Epidemiology* 221.

Becker A. B., 'Determinants of public support for same-sex marriage: Generational cohorts, social contact, and shifting attitudes' (2012) 24(4) *International Journal of Public Opinion Research* 524.

Becker G. S., 'Crime and punishment: An economic approach' (1968) 76(2) *Journal of Political Economy* 169.

Bell J. G. and Perry B., 'Outside looking in: The community impacts of anti-lesbian, gay, and bisexual hate crime' (2015) 62(1) *Journal of Homosexuality* 98.

Benjamin A. J. and Bushman B. J., 'The weapons priming effect' (2016) 12 *Current Opinion in Psychology* 45.

Benjamin A. J., Kepes S. and Bushman B. J., 'Effects of weapons on aggressive thoughts, angry feelings, hostile appraisals, and aggressive behavior: A meta-analytic review of the weapons effect literature' (2018) 22(4) *Personality and Social Psychology Review* 347.

Bennett E., 'Neuroscience and criminal law; Have we been getting it wrong for centuries and where do we go from here?' (2016) 85(2) *Fordham Law Review* 437.

Benson-Amram S., Heinen V. K., Dryer S. L. and Holekamp K. E., 'Numerical assessment and individual call discrimination by wild spotted hyaenas, *Crocuta crocuta*' (2011) 82(4) *Animal Behaviour* 743.

Berkowitz L. and LePage A., 'Weapons as aggression-eliciting stimuli' (1967) 7(2) *Journal of Personality and Psychology* 202.

Berman M. N., 'The normative functions of coercion claims' (2002) 8(1) *Legal Theory* 45.

Bernier A., Carlson S. M., Deschênes M. and Matte-Gagné C., 'Social factors in the development of early executive functioning: A closer look at the caregiving environment' (2012) 15(1) *Developmental Science* 12.

Birch D. J., 'The foresight saga: The biggest mistake of all?' (1988) (Jan) *Criminal Law Review* 4.

Birrell J. M. and Brown V. J., 'Medial frontal cortex mediates perceptual attentional set shifting in the rat' (2000) 20(11) *Journal of Neuroscience* 4320.

Blackmore S. and Troscianko E. T., *Consciousness: An Introduction* (3rd ed. Routledge 2018).

Blair C., 'Developmental science and executive function' (2016) 25(1) *Current Directions in Psychological Science* 3.

Blair C., 'Stress and the development of self-regulation in context' (2010) 4(3) *Child Development Perspectives* 181.

Blair C., Granger D. A., Kivlinghan K. T., Mills-Koonce R., Willoughby M., Greenberg M. T., Hibel L. C., Fortunato C. K. and Family Life Project Investigators, 'Maternal and child contributions to cortisol response to emotional arousal in young children from low-income, rural communities' (2008) 44(4) *Developmental Psychology* 1095.

Blair C., Granger D. A., Willoughby M., Mills-Koonce R., Cox M., Greenberg M. T., Kivlinghan K. T., Fortunato C. K. and Family Life Project Investigators, 'Salivary cortisol mediates effects of poverty and parenting on executive functions in early childhood' (2011) 82(6) *Child Development* 1970.

Blakemore S.-J. and Sirigu A., 'Action prediction in the cerebellum and in the parietal lobe' (2003) 153(2) *Experimental Brain Research* 239.

Blakemore S.-J., Wolpert D. and Frith C., 'Why can't you tickle yourself?' (2000) 11(11) *Neuroreport* R11.

Blouet C. and Schwartz G. J., 'Hypothalamic nutrient sensing in the control of energy homeostasis' (2010) 209(1) *Behavioural Brain Research* 1.

Bokura H., Yamaguchi S. and Kobayashi S., 'Electrophysiological correlates for response inhibition in a Go/NoGo task' (2001) 112(12) *Clinical Neurophysiology* 2224.

Bonn G. B., 'Re-conceptualizing free will for the 21st century: Acting independently with a limited role for consciousness' (2013) 4 *Frontiers in Psychology* 1.

Bornman E. and Mynhardt J. C., 'Social identity and intergroup contact with specific reference to the work situation' (1991) 117(4) *Genetic, Social, and General Psychology Monographs* 437.

Bösche W., 'Violent video games prime both aggressive and positive cognitions' (2010) 22(4) *Journal of Media Psychology* 139.

Botvinick M. M., Braver T. S., Barch D. M., Carter C. S. and Cohen J. D., 'Conflict monitoring and cognitive control' (2001) 108(3) *Psychological Review* 624.

Botvinick M. M., Cohen J. D. and Carter C. S., 'Conflict monitoring and anterior cingulate cortex: an update' (2004) 8(12) *Trends in Cognitive Sciences* 539.

Bourque C. W., 'Central mechanisms of osmosensation and systemic osmoregulation' (2008) 9(7) *Neuroscience* 519.

Boyd J. and Ingberman D. E., 'Do punitive damages promote deterrence?' (1999) 19(1) *International Review of Law and Economics* 47.

Braakmann N., 'The link between crime risk and property prices in England and Wales: Evidence from street-level data' (2016) 54(8) *Urban Studies* 1990.

Brasil-Neto J. P., Pascaul-Leone A., Valls-Solé J., Cohen L. G. and Hallett M., 'Focal transcranial magnetic stimulation and response bias in a forced-choice task' (1992) 55(10) *Journal of Neurology, Neurosurgery, and Psychiatry* 964.

Brass M. and Haggard P., 'To do or not to do: The neural signature of self-control' (2007) 27(34) *Journal of Neuroscience* 9141.

Brass M. and Haggard P., 'The what, when, whether model of intentional action' (2008) 14(4) *Neuroscientist* 319.

Brass M., Lynn M. T., Demanet J. and Rigoni D., 'Imaging volition: What the brain can tell us about the will' (2013) 229(3) *Experimental Brain Research* 301.

Brookbanks W., 'Compulsion and self-defence' (1990) 20(1) *Victoria University of Wellington Law Review* 95.

Brookes J. C., Hartoutsiou F., Horsfield A. P. and Stoneham A. M., 'Could humans recognise odor by phonon assisted tunnelling?' (2007) 98(3) *Physical Review Letters* 3.

Brown R., Croizet J.-C., Bohner G., Fournet M. and Payne A., 'Automatic category activation and social behaviour: The moderating role of prejudiced beliefs' (2003) 21(3) *Social Cognition* 167.

Buchanan T., 'Aggressive priming online: Facebook adverts can prime aggressive cognitions' (2015) 48 *Computers in Human Behavior* 323.

Bundesen C., Habekost T. and Kyllingsbaek S., 'A neural theory of visual attention: Bridging cognition and neurophysiology' (2005) 112(2) *Psychological Review* 291.

Burgess P. W., Scott S. K. and Frith C. D., 'The role of the rostral frontal cortex (area 10) in prospective memory: a lateral versus medial dissociation' (2003) 41(8) *Neuropsychologia* 906.

Burns J. M. and Swerdlow R. H., 'Right orbitofrontal tumor with pedophilia symptom and constructional apraxia sign' (2003) 60(3) *Archives of Neurology* 437.

Busemeyer J. R. and Townsend J. T., 'Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment' (1993) 100(3) *Psychological Review* 432.

Bushman B. J., 'Moderating role of trait aggressiveness in the effects of violent media on aggression' (1995) 69(5) *Journal of Personality and Social Psychology* 950.

Bushman B. J., 'Priming effects of media violence on the accessibility of aggressive constructs in memory' (1998) 24(5) *Personality and Social Psychology Bulletin* 537.

Bushman B. J. and Anderson C. A., 'Media violence and the American public: Scientific fact versus media misinformation' (2001) 56(6/7) *American Psychologist* 477.

Bushman B. J. and Baumeister R. F., 'Threatened egoism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence?' (1998) 75(1) *Journal of Personality and Social Psychology* 219.

Cabral H. O., Vinck M., Fouquet C., Pennartz C. M. A., Rondi-Reig L. and Battaglia F. P., 'Oscillatory dynamics and place field maps reflect hippocampal ensemble processing of sequence and place memory under NMDA receptor control' (2014) 81(2) *Neuron* 402.

Campbell K., 'The test of dishonesty in *R v Ghosh*' (1984) 43(2) *Cambridge Law Journal* 349.

Campbell-Meiklejohn D. K., Woolrich M. W., Passingham R. E. and Rogers R. D., 'Knowing when to stop: The brain mechanisms of chasing losses' (2008) 63(3) *Biological Psychiatry* 293.

Capella M. L., Hill R. P., Rapp J. M. and Kees J., 'The impact of violence against women in advertisements' (2010) 39(4) *Journal of Advertising* 37.

Carlson M., Marcus-Newhall A. and Miller N., 'Effects of situational aggression cues: A quantitative review' (1990) 58(4) *Journal of Personality and Social Psychology* 622.

Carlson S. M. and Moses L. J., 'Individual differences in inhibitory control and children's theory of mind' (2001) 72(4) *Child Development* 1032.

Carlson S. M. and Zelazo P. D., 'The value of control and the influence of values' (2011) 108(41) *Proceedings of the National Academy of Sciences* 16861.

Carpentier F. R. D., Northup C. T. and Parrott M. S., 'Revisiting media priming effects of sexual depictions: Replication, extension, and consideration of sexual depiction strength' (2014) 17(1) *Media Psychology* 34.

Carr M. F., Jadhav S. P. and Frank L. M., 'Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval' (2011) 14(2) *Nature Neuroscience* 147.

Carter C. S. and van Veen V., 'Anterior cingulate cortex and conflict detection: An update of theory and data' (2007) 7(4) *Cognitive, Affective, & Behavioural Neuroscience* 367.

Caruso G. D. and Morris S. G., 'Compatibilism and retributivist desert moral responsibility: On what is of central philosophical and practical importance' (2017) 82(4) *Erkenntnis* 837.

Carver C. S., Ganellen R. J., Froming W. J. and Chambers W., 'Modeling: An analysis in terms of category accessibility' (1983) 19(5) *Journal of Experimental Social Psychology* 403.

Casey B. J., Somerville L. H., Gotlib I. H., Ayduk O., Franklin N. T., Askren M. K., Jonides J., Berman M. G., Wilson N. L., Teslovich T., Glover G., Zayaz V., Mischel W. and Shoda Y., 'Behavioral and neural correlates of delay of gratification 40 years later' (2011) 108(36) *Proceedings of the National Academy of Sciences* 14998.

Caspers S., Zilles K., Laird A. R. and Eickhoff S. B., 'ALE meta-analysis of action observation and imitation in the human brain' (2010) 50(3) *NeuroImage* 1148.

Cesario J., 'Priming, replication, and the hardest science' (2014) 9(1) *Perspectives on Psychological Science* 40.

Cesario J. and Jonas K. J., 'Replicability and models of priming: What a resource computation framework can tell us about expectations of replicability' (2014) 32(Supp) *Social Cognition* 124.

Cesario J. and Navarrete C. D., 'Perceptual bias in threat distance: The critical roles of in-group support and target evaluations in defensive threat regulation' (2013) 5(1) *Social Psychological and Personality Science* 12.

Cesario J., Plaks J. E., Hagiwara N. and Navarrete C. D., 'The ecology of automaticity: How situational contingencies shape action semantics and social behavior' (2010) 21(9) *Psychological Science* 1311.

Cesario J., Plaks J. E. and Higgins E. T., 'Automatic social behavior as motivated preparation to interact' (2006) 90(6) *Journal of Personality and Social Psychology* 893.

Chalfin A. and McCrary J., 'Criminal deterrence: A review of the literature' (2017) 55(1) *Journal of Economic Literature* 5.

Chambon V., Wenke D., Fleming S. M., Prinz W. and Haggard P., 'An online neural substrate for sense of agency' (2013) 23(5) *Cerebral Cortex* 1031.

Chandler M. J., Sokol B. W. and Wainryb C., 'Beliefs about truth and beliefs about rightness' (2000) 71(1) *Child Development* 91.

Chang H.-B., 'Attitudes of Chinese students in the United States' (1973) 58(1) *Sociology and Social Research* 66.

Chantry A. D., '*R v Moloney* and the mental element in murder' (1985) 7(2) *Liverpool Law Review* 168.

Chapman C. S., Gallivan J. P., Wood D. K., Milne J. L., Culham J. C. and Goodale M. A., 'Short-term motor plasticity revealed in a visuomotor decision-making task' (2010) 214(1) *Behavioural Brain Research* 120.

Charness G. and Gneezy U., 'Strong evidence for gender differences in risk taking' (2012) 83(1) *Journal of Economic Behavior & Organization* 50.

Charron S. and Koechlin E., 'Divided representation of concurrent goals in the human frontal lobes' (2010) 328(5976) *Science* 360.

Chartrand T. L. and Bargh J. A., 'Automatic activation of impression formation and memorization goals: Nonconscious goal priming reproduces effect of explicit task instructions' (1996) 71(3) *Journal of Personality and Social Psychology* 464.

Chen E. Y., 'Impacts of "three strikes and you're out" on crime trends in California and throughout the United States' (2008) 24(4) *Journal of Contemporary Criminal Justice* 345.

Cheung T. T. L., Junghans A. F., Dijskterhuis G. B., Kroese F. M., Johansson P., Hall L. and de Ridder D. T. D., 'Consumers' choice-blindness to ingredient information' (2016) 106 *Appetite* 2.

Child J. J. and Sullivan G. R., 'When does the insanity defence apply? Some recent cases' (2014) 11 *Criminal Law Review* 788.

Christenson R., 'The political theory of persecution: Augustine and Hobbes' (1968) 12(3) *Midwest Journal of Political Science* 419.

Christopher R., 'Deterring retributivism: The injustice of "just" punishment' (2002) 96(3) *Northwestern University Law Review* 843.

Churchland P. S., 'Brain-base values' (2005) 93(4) *American Scientist* 356.

Cisek P., 'Cortical mechanisms of action selection: The affordance competition hypothesis' (2007) 362(1485) *Philosophical Transactions of the Royal Society: Biological Sciences* 1585.

Cisek P., 'Integrated neural processes for defining potential actions and deciding between them: A computational model' (2006) 26(38) *Journal of Neuroscience* 9761.

Cisek P., 'Making decisions through a distributed consensus' (2012) 22(6) *Current Opinion in Neurobiology* 927.

Cisek P. and Kalaska J. F., 'Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action' (2005) 45 (5) *Neuron* 801.

Cisek P. and Kalaska J. F., 'Neural mechanisms for interacting with a world full of action choices' (2010) 33(1) *Annual Review of Neuroscience* 269.

Cisek P. and Kalaska J. K., 'Simultaneous encoding of multiple potential reach directions in dorsal premotor cortex' (2002) 87(2) *Journal of Neurophysiology* 1149.

Claxton G., 'Why can't we tickle ourselves' (1975) 41(1) *Perceptual and Motor Skills* 335.

Claydon L., 'Reforming automatism and insanity: Neuroscience and claims of lack of capacity for control' (2015) 55(3) *Medicine, Science and the Law* 162.

Clément F., 'To trust or not to trust? Children's social epistemology' (2010) 1(4) *Review of Philosophy and Psychology* 531.

Clough J., 'Giving up the *Ghosh*: *Ivey (Appellant) v Genting Casinos (UK) Ltd trading as Crockfords (Respondent)*' (2018) 236 *Criminal Lawyer* 2.

Coffey G., 'Codifying the meaning of "intention" in the criminal law' (2009) 73(5) *Journal of Criminal Law* 394.

Cohen A. J. and Leckman J. F., 'Sensory phenomena associated with Gilles de la Tourette's syndrome' (1992) 53(9) *Journal of Clinical Psychiatry* 319.

Cohn A., Fehr E. and Maréchal M. A., 'Business culture and dishonesty in the banking industry' (2014) 516(7529) *Nature* 86.

Colvin E., 'Exculpatory defences in criminal law' (1990) 10(3) *Oxford Journal of Legal Studies* 381.

Cooney J. W. and Gazzaniga M. S., 'Neurological disorders and the structure of human consciousness' (2003) 7(4) *Trends in Cognitive Sciences* 161.

Copp D., '"Ought" implies "can" and the derivation of the principle of alternate possibilities' (2008) 68(1) *Analysis* 67.

Cos I., Bélanger N. and Cisek P., 'The influence of predicted arm biomechanics on decision making' (2011) 105(6) *Journal of Neurophysiology* 3022.

Cos I., Medleg F. and Cisek P., 'The modulatory influence of end-point controllability on decisions between actions' (2012) 108(6) *Journal of Neurophysiology* 1764.

Cottingham J., 'Varieties of retribution' (1979) 29(116) *Philosophical Quarterly* 238.

Coughlan S., 'The rise and fall of duress: How duress changed necessity before being excluded by self-defence' (2013) 39(1) *Queen's Law Journal* 83.

Cowan V., 'Reckless driving – Recklessness not limited to the subjective test – Extends to cases where D fails to give any thought to the possibility of risk – Direction to jury' (1992) (Nov) *Criminal Law Review* 814.

Cowley D., 'Defence of duress – The objective test' (1997) 61(2) *Journal of Criminal Law* 178.

Coyne S. M., Linder J. R., Nelson D. A. and Gentile D. A., '"Frenemies, Fraitors, and Mean-em-aitors": Priming effects of viewing physical and relational aggression in the media on women' (2012) 38(2) *Aggressive Behavior* 141.

Crittenden J. R. and Graybiel A. M., 'Basal ganglia disorders associated with imbalances in the striatal striosome and matrix compartments' (2011) 5 *Frontiers in Neuroanatomy* 1.

Crone T. S., 'The effect of nonconscious goal conflict on goal-related behavior' (2016) 3(3) *Psychology of Consciousness: Theory, Research and Practice* 284.

Crosby C., 'Gross negligence manslaughter revisited: Time for a change of direction?' (2020) 84(3) *Journal of Criminal Law* 228.

Crosby C., 'Recklessness – The continuing search for a definition' (2008) 72(4) *Journal of Criminal Law* 313.

Cui H. and Andersen R. A., 'Posterior parietal cortex encodes autonomously selected motor plans' (2007) 56(3) *Neuron* 552.

Cullen F. T., Jonson C. L. and Nagin D. S., 'Prisons do not reduce recidivism: The high cost of ignoring science' (2011) 91(3Supp) *The Prison Journal* 48S.

Cunningham S., 'Recklessness: Being reckless and acting recklessly' (2010) 21(3) *King's Law Journal* 445.

Cunnington R., Windischberger C., Deecke L. and Moser E., 'The preparation and readiness for voluntary movement: a high-field event-related fMRI study of the Bereitschafts-BOLD response' (2003) 20(1) *NeuroImage* 404.

Cunnington R., Windischberger C. and Moser E., 'Premovement activity of the pre-supplementary motor area and the readiness for action: Studies of time-resolved event-related functional MRI' (2005) 24(5-6) *Human Movements Science* 644.

Cushman F. and Young L., 'Patterns of moral judgment derive from nonmoral psychological representation' (2011) 35 (6) *Cognitive Science: A Multidisciplinary Journal* 1052.

Cushman F., Young L. and Hauser M., 'The role of conscious reasoning and intuition in moral judgment' (2006) 17(12) *Psychological Science* 1082.

Custers R. and Aarts H., 'Beyond priming effects: The role of positive affect and discrepancies in implicit processes of motivation and goal pursuit' (2005a) 16(1) *European Review of Social Psychology* 257.

Custers R. and Aarts H., 'Positive affect as implicit motivator: On the nonconscious operation of behavioral goals' (2005b) 89(2) *Journal of Personality and Social Psychology* 129.

Custers R. and Aarts H., 'In search of the nonconscious sources of goal pursuit: Accessibility and positive affective valence of the goal state' (2007) 43(2) *Journal of Experimental Social Psychology* 312.

Dashwood A., 'Logic and the Lords in *Majewski*' (1976) *Criminal Law Review* 532.

Davies J., Zhu H. and Brantley B., 'Sex appeals that appeal: Negative sexual self-schema as a moderator of the priming effects of sexual ads on accessibility' (2007) 29(2) *Journal of Current Issues & Research in Advertising* 79.

Davies M., 'Lawmakers, law lords and legal fault: Two tales from the Thames River bank: Sexual Offences Act 2003; R v G and another' (2004) 68(2) *Journal of Criminal Law* 130.

Debaere F., Wenderoth N., Sunaert S., van Hecke P. and Swinnen S. P., 'Internal vs external generation of movements: differential neural pathways involved in bimanual coordination performed in the presence or absence of augmented visual feedback' (2003) 19(3) *NeuroImage* 764.

DeBono A., Shariff A. F., Poole S. and Muraven M., 'Forgive us our trespasses: Priming a forgiving (but not a punishing) God increases unethical behavior' (2017) 9(Supp 1) *Psychology of Religion and Spirituality* S1.

Deecke L., 'Planning, preparation, execution, and imagery of volitional action' (1996) 3(2) *Cognitive Brain Research* 59.

Deecke L., Grözinger B. and Kornhuber H. H., 'Voluntary finger movement in man: Cerebral potentials and theory' (1976) 23(2) *Biological Cybernetics* 99.

Deecke L. and Kornhuber H. H., 'An electrical sign of participation of the mesial "supplementary" motor cortex in human voluntary finger movement' (1978) 159(2) *Brain Research* 473.

Deecke L., Scheid P. and Kornhuber H. H., 'Distribution of readiness potential, pre-motion positivity, and motor potential of the human cerebral cortex preceding voluntary finger movements' (1969) 7(2) *Experimental Brain Research* 158.

de Groot A. M. D., 'The range of automatic spreading activation in word priming' (1983) 22(4) *Journal of Verbal Learning and Verbal Behavior* 417.

Dehaene S., Artiges E., Naccache L., Martelli C., Viard A., Schürhoff F., Recasens C., Martinot M. L. P., Leboyer M. and Martinot J.-L., 'Conscious and subliminal conflicts in normal subjects and patients with schizophrenia: The role of the anterior cingulate' (2003) 100(23) *Proceedings of the National Academy of Sciences* 13722.

Dehaene S. and Naccache L., 'Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework' (2001) 79(1/2) *Cognition* 1.

Dehaene S., Naccache L., le Clec'H G., Koechlin E., Mueller M., Dehaene-Lambertz G., van de Moortele P.-F. and le Bihan D., 'Imaging unconscious semantic priming' (1998) 395(6702) *Nature* 597.

Deiber M.-P., Honda M., Ibañez V., Sadato N. and Hallett M., 'Mesial motor areas in self-initiated versus externally triggered movements explained with fMRI: Effect of movement type and rate' (1999) 81(6) *Journal of Neurophysiology* 3065.

DeMarree K. G. and Loersch C., 'Who am I and who are you? Priming and the influence of self versus other focused attention' (2009) 45 *Journal of Experimental Social Psychology* 440.

DeMarree K. G., Loersch C., Briñol P., Petty R. E., Payne B. K. and Rucker D. D., 'From primed construct to motivated behavior: Validation processes in goal pursuit' (2012) 38(12) *Personality and Social Psychology Bulletin* 1659.

DeMarree K. G., Wheeler S. C. and Petty R. E., 'Priming a new identity: Self-monitoring moderates the effects of nonself primes on self-judgments and behavior' (2005) 89(5) *Journal of Personality and Social Psychology* 657.

Demetriou K., 'The soft-line solution to Pereboom's Four-Case Argument' (2010) 88(4) *Australasian Journal of Philosophy* 595.

Dennett D. C., 'How to study human consciousness empirically or nothing comes to mind' (1982) 53(2) *Matters of the Mind* 159.

Dennett D. C., 'The self as a responding – and responsible – artifact' (2003) 1001(1) *Annals of the New York Academy of Sciences* 39.

Dent N. and Kervick Á., 'Ghosh: A change in direction?' (2016) 8 *Criminal Law Review* 553.

Denson T. F., Capper M. M., Oaten M., Friese M. and Schofield T. P., 'Self-control training decreases aggression in response to provocation in aggressive individuals' (2011) 45(2) *Journal of Research in Personality* 252.

Desforges D. M., Lord C. G., Ramsey S. L., Mason J. A., van Leeuwen M. D., West S. C. and Leper M. R., 'Effects of structured cooperative contact on changing negative attitudes toward stigmatized social groups' (1991) 60(4) *Journal of Personality and Social Psychology* 531.

Desmurget M., Reilly K. T., Richard N., Szathmari A., Mottolese C. and Sirigu A., 'Movement intention after parietal cortex stimulation in humans' (2009) 324(5928) *Science* 811.

Desmurget M. and Sirigu A., 'A parietal-premotor network for movement intention and motor awareness' (2009) 13(10) *Trends in Cognitive Science* 411.

Devinsky O., Morrell M. J. and Vogt B. A., 'Contributions of anterior cingulate cortex to behaviour' (1995) 118(1) *Brain* 279.

DeWall C. N., Baumeister R. F., Stillman T. F. and Gailliot M. T., 'Violence restrained: Effects of self-regulation and its depletion on aggression' (2007) 43(1) *Journal of Experimental Social Psychology* 62.

DeWall C. N., Finkel E. J. and Denson T. F., 'Self-control inhibits aggression' (2011) 5(7) *Social and Personality Psychology Compass* 458.

Diamond A., 'Executive functions' (2013) 64(1) *Annual Review of Psychology* 135.

Diamond A., Barnett W. S., Thomas J. and Munro S., 'Preschool program improves cognitive control' (2007) 318(5855) *Science* 1387.

Dias R., Robbins T. W., Roberts A. C., 'Dissociation in prefrontal cortex of affective and attentional shifts' (1996) 380(6569) *Nature* 69.

Dickson S. and Stuart-Cole E., 'Mentally relevant? When is a loss of control attributable to a mental condition?' (2018) 82(2) *Journal of Criminal Law* 117.

Dijksterhuis A., Aarts H., Bargh J. A. and van Knippenberg A., 'On the relation between associative strength and automatic behavior' (2000) 36(5) *Journal of Experimental Social Psychology* 531.

Dik G. and Aarts H., 'Behavioral cues to others' motivation and goal pursuits: The perception of effort facilitates goal inference and contagion' (2007) 43(5) *Journal of Experimental Social Psychology* 727.

Doebel S., Michaelson L. E. and Munakata Y., 'Good things come to those who wait: Delaying gratification likely does matter for later achievement (A commentary on Watts, Duncan, & Quan, 2018)' (2019) *Psychological Science* 1.

Dolinko D., 'Some thoughts about retributivism' (1991) 101(3) *Ethics* 537.

Dölling D., Entorf H., Hermann D. and Rupp T., 'Is deterrence effective? Results of a meta-analysis of punishment' (2009) 15(1-2) *European Journal on Criminal Policy and Research* 201.

Donnerstein E. and Linz D., 'Mass media sexual violence and male viewers' (1986) 29(5) *American Berhavioral Scientist* 601.

Doob A. N. and Webster C. M., 'Sentence severity and crime: Accepting the null hypothesis' (2003) 30(1) *Crime and Justice* 143.

Dragori G. and Tonegawa S., 'Preplay of future place cell sequences by hippocampal cellular assemblies' (2010) 469(7330) *Nature* 397.

Driver J. and Mattingley J. B., 'Parietal neglect and visual awareness' (1998) 1(1) *Nature Neuroscience* 17.

Dubreuil B., 'Paleolithic public goods games: why human culture and cooperation did not evolve in one step' (2009) 25(1) *Biology & Philosophy* 53.

Duckworth A. L., Tsukayama E. and Greier A. B., 'Self-controlled children stay leaner in the transition to adolescence' (2010) 54(2) *Appetite* 304.

Duff R. A., '*Caldwell* and *Lawrence*: The retreat from subjectivism' (1983) 3(1) *Oxford Journal of Legal Studies* 77.

Duff R. A., 'The obscure intentions of the House of Lords' (1986) (Dec) *Criminal Law Review* 771.

Duggal H. S. and Nizamie S. H., 'Bereitschaftspotential in tic disorders: A preliminary observation' (2002) 50(4) *Neurology India* 487.

Dunbar R. I. M., 'The social brain hypothesis' (1998) 6(5) *Evolutionary Anthropology: Issues, News, and Reviews* 178.

Dunbar R. I. M. and Shultz S., 'Evolution in the social brain' (2007) 317(5843) *Science* 1344.

Dyson M., 'Some aspects of the vibration theory of odor' (1928) 19 *Perfumery and Essential Oil Record* 456.

Dyson M. and Jarvis P., 'Poison Ivey or herbal tea leaf?' (2018) 134(Apr) *Law Quarterly Review* 198.

Ebrahim I., Fenwick P., Marks R. and Peacock K. W., 'Violence, sleepwalking and the criminal law: Part 1: The medical aspects' (2005) (Aug) *Criminal Law Review* 601.

Ebrahim I. and Fenwick P., 'Sleep related automatism and the law' (2008) 48(2) *Medicine, Science and the Law* 124.

Edwards J. Ll. J., 'The criminal degrees of knowledge' (1954) 17(4) *Modern Law Review* 294.

Edwards K. and von Hippel W., 'Hearts and minds: The priority of affective versus cognitive factors in person perception' (1995) 21(10) *Personality and Social Psychology Bulletin* 996.

Eichenbaum H., Dudchenko P., Wood E., Shapiro M. and Tanila H., 'The hippocampus, memory, and place cells: Is it spatial memory or a memory space?' (1999) 23(2) *Neuron* 209.

Eigsti I.-M., Zayas V., Mischel W., Shoda Y., Ayduk O., Dadlani M. B., Davidson M. C., Aber J. L. and Casey B. J., 'Predicting cognitive control from preschool to late adolescence and young adulthood' (2006) 17(6) *Psychological Science* 478.

Eimer M., 'Facilitatory and inhibitory effects of masked prime stimuli on motor activation and behavioural performance' (1999) 101(2/3) *Acta Psychologica* 293.

Eimer M. and Schlaghecken F., 'Effects of masked stimuli on motor activation: Behavioral and electrophysical evidence' (1998) 24(6) *Journal of Experimental Psychology* 1737.

Eimer M. and Schlaghecken F., 'Response facilitation and inhibition in subliminal priming' (2003) 64(1/2) *Biological Psychology* 7.

Eimeren T., Wolbers T., Münchau A., Büchel A., Weiller C. and Siebner H. R., 'Implementation of visuospatial cues in response selection' (2006) 29(1) *NeuroImage* 286.

Elkington A., 'The historical development of duress and the unfounded result of denying duress as a defence to murder' (2022) 0(0) [online] *Journal of Criminal Law* 1.

Engle E., 'The history of the general principle of proportionality: An overview' (2012) 10(1) *Dartmouth Law Journal* 1.

Evans J. St. B. T., 'Logic and human reasoning: An assessment of the deduction paradigm' (2002) 128(6) *Psychological Bulletin* 978.

Evans J. St. B. T. and Wason P. C., 'Rationalization in a reasoning task' (1976) 67(4) *British Journal of Psychology* 479.

Faber T. W. and Jonas K. J., 'Perception in a social context: Attention for response-functional means' (2013) 31(2) *Social Cognition* 301.

Falk A. and Fischbacher U., '"Crime" in the lab – Detecting social interaction' (2002) 46(4/5) *European Economic Review* 859.

Falk A., Kosse F. and Pinger P., 'Re-revisiting the marshmallow test: A direct comparison of studies by Shoda, Mischel and Peake (1990) and Watts, Duncan, and Quan (2018)' (2019) *Psychological Science* 1.

Falkenstein M., Hoormann J. and Hohnsbein J., 'ERP components in Go/NoGo tasks and their relation to inhibition' (1999) 101(2/3) *Acta Psychologica* 267.

Farrer C., Bouchereau M., Jeannerod M. and Franck N., 'Effect of distorted visual feedback on the sense of agency' (2008) 19(1-2) *Behavioural Neurology* 53.

Farrer C., Franck N., Georgieff N., Frith C., Decety J. and Jeannerod M., 'Modulating the experience of agency: A positron emission tomography study' (2003) 18(2) *Neuroimage* 324.

Farrer C. and Frith C., 'Experiencing oneself vs another person as being the cause of an action: The neural correlates of the experience of agency' (2002) 15(3) *Neuroimage* 596.

Fazel S., Hayes A. J., Bartellas K., Clerici M. and Trestman R., 'The mental health of prisoners: A review of prevalence, adverse outcomes and interventions' (2016) 3(9) *Lancet Psychiatry* 871.

Fazel S. and Seewald K., 'Severe mental illness in 33,588 prisoners worldwide: Systematic review and meta-regression analysis' (2012) 200(5) *British Journal of Psychiatry* 364.

Fazio R. H. and Olson M. A., 'Implicit measures in social cognition research: Their meaning and use' (2003) 54(1) *Annual Review of Psychology* 297.

Feess E., Schildberg-Hörisch H., Schramm M. and Wohlschlegel A., 'The impact of fine size and uncertainty on punishment and deterrence: Theory and evidence from the laboratory' (2018) 149(1) *Journal of Economic Behavior and Organization* 58.

Feinberg I., 'Efference copy and corollary discharge: Implications for thinking and its disorders' (1978) 4(4) *Schizophrenia Bulletin* 636.

Feinberg J., 'The expressive function of punishment' (1965) 49(3) *The Monist* 397.

Feltz A., 'Pereboom and premises: Asking the right questions in the experimental philosophy of free will' (2013) 22(1) *Consciousness and Cognition* 53.

Ferguson M. J., Bargh J. A. and Nayak D. A., 'After-affects: How automatic evaluations influence the interpretation of subsequent, unrelated stimuli' (2005) 41(2) *Journal of Experimental Social Psychology* 182.

Fessler D. M. T. and Holbrook C., 'Friends shrink foes: The presence of comrades decreases the envisioned physical formidability of an opponent' (2013) 24(5) *Psychological Science* 797.

Fiehler K., Bannert M. M., Bischoff M., Blecker C., Stark R., Vaitl D., Franz V. H. and Rösler F., 'Working memory maintenance of grasp-target information in the human posterior parietal cortex' (2010) 54(3) *NeuroImage* 2401.

Field S. and Lynn M., 'Capacity, recklessness and the House of Lords' (1993) (Feb) *Criminal Law Review* 127.

Field S. and Lynn M., 'The capacity for recklessness' (1992) 12(1) *Legal Studies* 74.

Filbey F. M. and DeWitt S. K., 'Cannabis cue-elicited craving and the reward neurocircuitry' (2012) 38(1) *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 30.

Filbey F. M., Schacht J. P., Myers U. S., Chavez R. S. and Hutchinson K. E., 'Marijuana craving in the brain' (2009) 106(31) *Proceedings of the National Academy of Sciences* 13016.

Filevich E., Kühn S. and Haggard P., 'Intentional inhibition in human action: The power of "no"' (2012) 36(4) *Neuroscience & Biobehavioral Reviews* 1107.

Filevich E., Kühn S. and Haggard P., 'There is no free won't: Antecedent brain activity predicts decisions to inhibit' (2013) 8(2) *PLoS ONE* e53053.

Filimon F., 'Human cortical control of hand movements: parietofrontal networks for reaching, grasping, and pointing' (2010) 16(4) *Neuroscientist* 388.

Filimon F., Nelson J. D., Huang R.-S. and Soreno M. I., 'Multiple parietal reach regions in humans: Cortical representations for visual and proprioceptive feedback during on-line reaching' (2009) 29(9) *Journal of Neuroscience* 2961.

Fine C., 'Is the emotional dog wagging its rational tail, or chasing it?' (2006) 9(1) *Philosophical Explorations* 83.

Finkel E. J., DeWall C. N., Slotter E. B., Oaten M. and Foshee V. A., 'Self-regulatory failure and intimate partner violence perpetration' (2009) 97(3) *Journal of Personality and Social Psychology* 483.

Finkel N. J., Liss M. B. and Moran V. R., 'Equal of proportional justice for accessories? Children's pearls of proportionate wisdom' (1997) 18(2) *Journal of Applied Developmental Psychology* 229.

Fischer P., Kastenmüller A. and Greitmeyer T., 'Media violence and the self: The impact of personalized gaming characters in aggressive video games on aggressive behavior' (2009) 46(1) *Journal of Experimental Social Psychology* 192.

Fishbach A. and Labroo A. A., 'Be better or be merry: How mood affects self-control' (2007) 93(2) *Journal of Personality and Social Psychology* 158.

Fiske S. T. and Neuberg S. L., 'A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation' (1990) 23 *Advances in Experimental Psychology* 1.

Flanagan-Cato L. M., 'Sex differences in the neural circuit that mediates female sexual receptivity' (2011) 32(2) *Frontiers in Neuroendocrinology* 124.

Fogassi L., Ferrari P. F., Gesierich B., Rozzi S., Chersi F. and Rizzolatti G., 'Parietal lobe: From action organization to intention understanding' (2005) 308(5722) *Science* 662.

Ford T. E. and Kruglanski A. W., 'Effects of epistemic motivations on the use of accessible constructs in social judgment' (1995) 21(9) *Personality and Social Psychology Bulletin* 950.

Forgas J. P. and Moylan S., 'After the movies: Transient mood and social judgments' (1987) 13(4) *Personality and Social Psychology Bulletin* 467.

Fowler C. A., Wolford G., Slade R. and Tassinary L., 'Lexical access with and without awareness' (1981) 110(3) *Journal of Experimental Psychology* 341.

Francis L. A. and Susman E. J., 'Self-regulation and rapid weight gain in children from age 3 to 12 years' (2009) 163(4) *Archives of Pediatrics and Adolescent Medicine* 297.

Franco M. I, Turin L., Mershin A. and Skoulakis E. M. C., 'Molecular vibration-sensing component in *Drosophila melanogaster* olfaction' (2011) 108(9) *Proceedings of the National Academy of Sciences* 3797.

Frankfurt H. G., 'Alternate possibilities and moral responsibility' (1969) 66(23) *Journal of Philosophy* 829.

Frankfurt H. G., 'Freedom of the will and the concept of a person' (1971) 68(1) *Journal of Philosophy* 5.

Franklin D. W. and Wolpert D. M., 'Computational mechanisms of sensorimotor control' (2011) 72(3) *Neuron* 425.

Frey S. H., Vinton D., Norlund R. and Grafton S. T., 'Cortical topography of human anterior intraparietal cortex active during visually guided grasping' (2005) 23(2-3) *Cognitive Brain Research* 397.

Fried I., Katz A., McCarthy G., Sass K. J., Williamson P., Spencer S. S. and Spencer D. D., 'Functional organization of human supplementary motor cortex studied by electrical stimulation' (1991) 11(11) *Journal of Neuroscience* 3656.

Fried I., Mukamel R. and Kreiman G., 'Internally generated preactivation of single neurons in human medial frontal cortex predicts volition' (2011) 69(3) *Neuron* 548.

Frodl T. and O'Keane V., 'How does the brain deal with cumulative stress? A review with focus on developmental stress, HPA axis function and hippocampal structure in humans' (2013) 52(1) *Neurobiology of Disease* 24.

Fruchtman E., 'Recklessness and the limits of *mens rea*: Beyond orthodox subjectivism: Part I' (1987a) 29(3) *Criminal Law Quarterly* 315.

Fruchtman E., 'Recklessness and the limits of *mens rea*: Beyond orthodox subjectivism: Part II' (1987b) 29(4) *Criminal Law Quarterly* 421.

Gaal S., de Lange F. P. and Cohen M. X., 'The role of consciousness in cognitive control and decision making' (2012) 6(121) *Frontiers in Human Neuroscience* 1.

Gaal S., Ridderinkhof R., Scholte H. S. and Lamme V. A. F., 'Unconscious activation of the prefrontal no-go network' (2010) 30(11) *Journal of Neuroscience* 4143.

Gaal S., Ridderinkhof R., van den Wildenberg W. P. M. and Lamme V. A. F., 'Dissociating consciousness from inhibitory control: Evidence for unconsciously triggered response inhibition in the stop-signal task' (2009) 35(4) *Journal of Experimental Psychology* 1129.

Gądek-Michalska A., Spyrka J., Rachwalska P., Tadeusz J. and Bugajsk J., 'Influence of chronic stress on brain corticosteroid receptors and HPA axis activity' (2013) 65(5) *Pharmacological Reports* 1163.

Gaes G. G. and Camp S. D., 'Unintended consequences: Experimental evidence for the criminogenic effect of prison security level placement on post-release recidivism' (2009) 5(2) *Journal of Experimental Criminology* 139.

Gailliot M. T. and Baumeister R. F., 'Self-regulation and sexual restraint: Dispositionally and temporarily poor self-regulatory abilities contribute to failures at restraining sexual behavior' (2007) 33(2) *Personality and Social Psychology Bulletin* 173.

Gailliot M. T., Plant E. A., Butz D. A. and Baumeister R. F., 'Increasing self-regulatory strength can reduce the depleting effect of suppressing stereotypes' (2007) 33(2) *Personality and Social Psychology Bulletin* 281.

Galli M., 'Oh my *Ghosh*: Supreme Court redefines test for dishonesty in *Ivey v Genting Casinos*' (2018) 29(2) *Entertainment Law Review* 55.

Gallivan J. P., Barton K. S., Chapman C. S., Wolpert D. M. and Flanagan J. R., 'Action plan co-optimization reveals the parallel encoding of competing reach movements' (2015) 6 *Nature Communications* 7428

Gallivan J. P., Bowman N. A. R., Chapman C. S., Wolpert D. M. and Flanagan J. R., 'The sequential encoding of competing action goals involves dynamic restructuring of motor plans in working memory' (2016) 115(6) *Journal of Neurophysiology* 3113.

Gallivan J. P., Chapman C. S., Wood D. K., Milne J. L., Ansari D., Culham J. C. and Goodale M. A., 'One to four, and nothing more: Nonconscious parallel individuation of objects during action planning' (2011) 22(6) *Psychological Science* 803.

Gallivan J. P., McLean D. A., Flanagan J. R. and Culham J. C., 'Where one hand meets the other: limb-specific and action-dependent movement plans decoded from preparatory signals in single human frontoparietal brain areas' (2013) 33(5) *Journal of Neuroscience* 1991.

Gallivan J. P., McLean D. A., Valyear K. F., Pettypiece C. E. and Culham J. C., 'Decoding action intentions from preparatory brain activity in human parieto-frontal networks' (2011) 31(26) *Journal of Neuroscience* 9599.

Gallivan J. P., Stewart B. M., Baugh L. A., Wolpert D. M. and Flanagan J. R., 'Rapid automatic motor encoding of competing reach options' (2017) 18(7) *Cell Reports* 1619.

Gallo I. S., Keil A., McCulloch K. C., Rockstroh B. and Gollwitzer P. M., 'Strategic automation of emotion regulation' (2009) 96(1) *Journal of Personality and Social Psychology* 11.

Gollwitzer P. M., 'Implementation intentions: Strong effects of simple plans' (1999) 54(7) *American Psychologist* 493.

Gantman A. P., Adriaanse M. A., Gollwitzer P. M. and Oettingen G., 'Why did I do that? Explaining actions activated outside of awareness' (2017) 24(5) *Psychonomic Bulletin & Review* 1563.

Gardner J., 'Rationality and the rule of law in offences against the person' (1994) 53(3) *Cambridge Law Journal* 502.

Gardner M. R., 'The *mens rea* enigma: Observations on the role of motive in the criminal law past and present' (1993) 3 *Utah Law Review* 635.

Gardner S., 'The importance of *Majewski*' (1994) 14(2) *Oxford Journal of Legal Studies* 279.

Gazzaniga M. S., Bogen J. E. and Sperry R. W., 'Observations on visual perception after disconnexion of the cerebral hemispheres in man' (1965) 88(2) *Brain: A Journal of Neurology* 221.

Gazzaniga M. S., Bogen J. E. and Sperry R. W., 'Some functional effects of sectioning the cerebral commissures in man' (1962) 48(10) *Proceedings of the National Academy of Sciences* 1765.

Gazzaniga M. S. and Sperry R. W., 'Language after section of the cerebral commissures' (1967) *Brain: A Journal of Neurology* 131.

Gentsch A., Schütz-Bosbach S., Endrass T. and Kathmann, N., 'Dysfunctional forward model mechanisms and aberrant sense of agency in obsessive-compulsive disorder' (2012) 71(7) *Biological Psychiatry* 652.

Gerritzen B. C. and Kirchgässner G., 'Facts or ideology: What determines the results of econometric estimates of the deterrence effect of death penalty? A meta-analysis' (2016) 4(6) *Open Journal of Social Sciences* 178.

Gibbons S., 'The costs of urban property crime' (2004) 114(499) *The Economic Journal* F441.

Gibson M., 'Diminished responsibility in *Golds* and beyond: Insights and implications' (2017) 7 *Criminal Law Review* 543.

Gilbert D. T., Krull D. S. and Pelham B. W., 'Of thoughts unspoken: Social interference and the self-regulation of behavior' (1988) 55(5) *Journal of Personality and Social Psychology* 685.

Gilbert S. J. and Burgess P. W., 'Executive function' (2008) 18(3) *Current Biology* R110.

Giles M., 'Judicial law-making in the criminal courts: the case of marital rape' (1992) (Jun) *Criminal Law Review* 407.

Gino F. and Ariely D., 'The dark side of creativity: Original thinkers can be more dishonest' (2012) 102(3) *Journal of Personality and Social Psychology* 445.

Gino F. and Bazerman M. H., 'When misconduct goes unnoticed: The acceptability of gradual erosion in others' unethical behavior' (2009) 45(4) *Journal of Experimental Social Psychology* 708.

Gino F. and Pierce L., 'The abundance effect: Unethical behavior in the presence of wealth' (2009) 109(2) *Organizational Behavior and Human Decision Processes* 142.

Gino F., Schweitzer M. E., Mead N. L. and Ariely D., 'Unable to resist temptation: How self-control depletion promotes unethical behavior' (2011) 115(2) *Organizational Behavior and Human Decision Processes* 191.

Glimcher P. W., 'Indeterminacy in brain and behavior' (2005) 56(1) *Annual Review of Psychology* 25.

Gneezy U. and Potters J., 'An experiment on risk taking and evaluation periods' (1997) 112(2) *Quarterly Journal of Economics* 631.

Goff L., 'The mental element in the crime of murder' (1987) 22(1) *Israel Law Review* 1.

Gordon G. H., 'Subjective and objective *mens rea*' (1974) 17 *Criminal Law Review* 355.

Gordon N. S. and Fondacaro M. R., 'Rethinking the voluntary act requirement: Implications from neuroscience and behavioral science research' (2018) 36(4) *Behavioral Sciences & the Law* 426.

Gowen E. and Hamilton A., 'Motor abilities in autism: A review using a computational context' (2013) 43(2) *Journal of Autism and Developmental Disorders* 323.

Graf P., Mandler G. and Haden P. E., 'Simulating amnesiac symptoms in normal subjects' (1982) 218(4578) *Science* 1243.

Grand S. and Segal S. J., 'Recovery in the absence of recall: An investigation of color-word interference' (1966) 72(1) *Journal of Experimental Psychology* 138.

Graybiel A. M., 'The basal ganglia and chunking of action repertoires' (1998) 70(1-2) *Neurobiology of Learning and Memory* 119.

Green S. P., 'The universal grammar of criminal law' (2000) 98(6) *Michigan Law Review* 2104.

Greenawalt K., 'The perplexing borders of justification and excuse' (1984) 84(8) *Columbia Law Review* 1897.

Greene J. and Cohen J., 'For the law, neuroscience changes nothing and everything' (2004) 359(1451) *Philosophical Transactions of the Royal Society: Biological Sciences* 1775.

Greenwald A. G. and Banaji M. R., 'Implicit social cognition: Attitudes, self-esteem, and stereotypes' (1995) 102(1) *Psychological Review* 4.

Greenwald A. G. and Farnham S., 'Using the implicit association test to measure self-esteem and self-concept' (2000) 79(6) *Journal of Personality and Social Psychology* 1022.

Greenwald A. G. and Lai C. K., 'Implicit social cognition' (2020) 71(25) *Annual Review of Psychology* 1.

Grent-'t-Jong T., Oostenveld R.., Medendorp W. P. and Praamstra P., 'Separating visual and motor components of motor cortex activation for multiple reach targets: A visuomotor adaptation study' (2015) 35(45) *Journal of Neuroscience* 15135.

Grèzes J. and Decety J., 'Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis' (2000) 12(1) *Human Brain Mapping* 1.

Griew E. J., 'Dishonesty: The objections to *Feely* and *Ghosh*' (1985) *Criminal Law Review* 341.

Griffiths M., 'Violent video games and aggression: A review of the literature' (1999) 4(2) *Aggression and Violent Behavior* 203.

Grill-Spector K. and Malach R., 'fMRI-adaptation: a tool for studying the functional properties of human cortical neurons' (2001) 107(1-3) *Acta Psychologica* 293.

Grim P. F., Kohlberg L. and White S. H., 'Some relations between conscience and attentional processes' (1968) 8(3) *Journal of Personality and Social Psychology* 239.

Grush R., 'The emulation theory of representation: motor control, imagery, and perception' (2004) 27(3) *Behavioral and Brain Sciences* 377.

Haggard P., 'Conscious intention and motor cognition' (2005) 9(6) *Trends in Cognitive Sciences* 290.

Haggard P., 'Human volition: towards a neuroscience of will' (2008) 9(12) *Nature Reviews Neuroscience* 934.

Haggard P., 'Sense of agency in the human brain' (2017) 18(4) *Nature Review Neuroscience* 196.

Haggard P. and Clark S., 'Intentional action: Conscious experience and neural prediction' (2003) 12(4) *Consciousness and Cognition* 695.

Haggard P., Clark S. and Kalogeras J., 'Voluntary action and conscious awareness' (2002) 5(4) *Nature Neuroscience* 382.

Haggard P. and Eimer M., 'On the relation between brain potentials and the awareness of voluntary movements' (1999) 126(1) *Experimental Brain Research* 128.

Haggard P., Martin F., Taylor-Clarke M., Jeanerod M. and Franck N., 'Awareness of action in schizophrenia' (2003) 14(7) *NeuroReport* 1081.

Hagger M. S., Wood C., Stiff C. and Chatzisarantis N. L. D., 'Ego depletion and the strength model of self-control: A meta-analysis' (2010) 136(4) *Psychological Bulletin* 495.

Haidt J., 'The emotional dog and its rational tail: A social intuitionist approach to moral judgment' (2001) 108(4) *Psychological Review* 814.

Haidt J., 'The new synthesis in moral psychology' (2007) 316(5827) *Science* 998.

Haidt J., Björklund F. and Murphy S., 'Moral dumbfounding: When intuition finds no reason' (2000) 1(2) *Lund Psychological Reports* 29.

Haidt J. and Hersh M. A., 'Sexual morality: The cultures and emotions of conservatives and liberals' (2001) 31(1) *Journal of Applied Social Psychology* 191.

Haidt J., Koller S. H. and Dias M. G., 'Affect, culture, and morality, or is it wrong to eat your dog?' (1993) 65(4) *Journal of Personality and Social Psychology* 613.

Haith A. M., Huberdeau D. M. and Krakauer J. W., 'Hedging your bets: Intermediate movements as optimal behavior in the context of an incomplete decision' (2015) 11(3) *PLOS Computational Biology* e1004171.

Haji I. and Cuypers S. E., 'Hard- and soft-line responses to Pereboom's four-case manipulation argument' (2006) 21(4) *Acta Analytica* 19.

Hale B., 'Dishonesty' (2019) 48(1-2) *Common Law World Review* 5.

Hall L. and Johansson P., 'Choice blindness: You don't know what you want' (2009) 2704 *New Scientist* 26.

Hall L., Johansson P. and Strandberg T., 'Lifting the veil of morality: Choice blindness and attitude reversals on a self-transforming survey' (2012) 7(9) *PLoS ONE* e45457.

Hall L., Johansson P., Tärning B., Sikström S. and Deutgen T., 'Magic at the marketplace: Choice blindness for the taste of jam and the smell of tea' (2010) 117(1) *Cognition* 54.

Hall L., Strandberg T., Pärnamets P., Lind A., Tärning B. and Johansson P., 'How the polls can be both spot on and dead wrong: Using choice blindness to shift political attitudes and voter intentions' (2013) 8(4) *PLoS ONE* e60554.

Halpin A. K. W., 'Intended consequences and unintentional fallacies' (1987) 7(1) *Oxford Journal of Legal Studies* 104.

Halpin A. K. W., 'The test for dishonesty' (1996) (May) *Criminal Law Review* 283.

Hameroff S. R. and Penrose R., 'Consciousness in the universe: A review of the "Orch OR" theory' (2014) 11(1) *Physics of Life Reviews* 39.

Hamilton A. F. de C. and Grafton S. T., 'Goal representation in human anterior intraparietal sulcus' (2006) 26(4) *Journal of Neuroscience* 1133.

Hamilton D. L., Katz L. B. and Leirer V. O., 'Cognitive representation of personality impressions: Organizational processes in first impression formation' (1980b) 39(6) *Journal of Personality and Social Psychology* 1050.

Hammond S. I., Müller U., Carpendale J. I. M., Bibok M. B. and Liebermann-Finestone D. P., 'The effects of parental scaffolding on preschoolers' executive function' (2012) 48(1) *Developmental Psychology* 271.

Hansen C. H. and Krygowski W., 'Arousal-augmented priming effects: Rock music videos and sex object schemas' (1994) 21(1) *Communication Research* 24.

Hardwick R. M., Rottschy C., Miall R. C. and Eickholl S. B., 'A quantitative meta-analysis and review of motor learning in the human brain' (2013) 67 *NeuroImage* 283.

Hare T., Tottenham N., Davidson M. C., Glover G. H. and Casey B. J., 'Contributions of amygdala and striatal activity in emotion regulation' (2005) 57(6) *Biological Psychiatry* 624.

Hartstra E., Waszak F. and Brass M., 'The implementation of verbal instructions: Dissociating motor preparation from the formation of stimulus-response associations' (2012) 63(3) *NeuroImage* 1143.

Hassin R. R., Aarts H. and Ferguson M. J., 'Automatic goal inferences' (2005) 41(2) *Journal of Experimental Social Psychology* 129.

Hassin R. R., Bargh J. A. and Zimerman S., 'Automatic and flexible: The case of non-conscious goal pursuit' (2009) 27(1) *Social Cognition* 20.

Hastie R. and Kumar P. A., 'Person memory: Personality traits as organizing principles in memory for behaviours' (1979) 37(1) *Journal of Personality and Social Psychology* 25.

Hathaway M., 'The moral significance of the insanity defence' (2009) 73(4) *Journal of Criminal Law* 310.

Hauser M., Cushman F., Young L., Jin R. K.-X. and Mikhail J., 'A dissociation between moral judgments and justifications' (2007) 22(1) *Mind & Language* 1.

Hauser M., Knoblich G., Repp B. H., Lautenschlager M., Gallinat J., Heinz A. and Voss M., 'Altered sense of agency in schizophrenia and the putative psychotic prodrome' (2011) 186(2-3) *Psychiatry Research* 170.

Hazan C. and Shaver P., 'Romantic love conceptualized as an attachment process' (1987) 52(3) *Journal of Personality and Social Psychology* 511.

Heilig M., Epstein D. H., Nader M., A. and Shaham Y., 'Time to connect: Bringing social context into addiction neuroscience' (2016) 17(9) *Nature Reviews Neuroscience* 592.

Herek G. M. and Capitanio J. P., '"Some of my best friends": Intergroup contact, concealable stigma, and heterosexuals' attitudes toward gay men and lesbians' (1996) 22(4) *Personality and Social Psychology Bulletin* 412.

Herring J. and Palser E., 'The duty of care in gross negligence manslaughter' (2007) (Jan) *Criminal Law Review* 24.

Higgins E. T., 'The aboutness principle: A pervasive influence on human inference' (1998) 16(1) *Social Cognition* 173.

Higgins E. T., Rholes W. S. and Jones C. R., 'Category accessibility and impression formation' (1977) 13(2) *Journal of Experimental Social Psychology* 141.

Histed M. H. and Miller E. K., 'Microstimulation of frontal cortex can reorder a remembered spatial sequence' (2006) 4(5) *PLoS Biology* e134.

Hitchler W. H., 'Necessity as a defence in criminal cases' (1929) 33(3) *Dickinson Law Review* 138.

Hok V., Save E., Lenck-Santini P.-P. and Poucet B., 'Coding for spatial goals in the prelimbic/infralimbic area of the rat frontal cortex' (2005) 102(12) *Proceedings of the National Academy of Sciences* 4602.

Hoffstaedter F., Grefkes C., Zilles K. and Eickhoff S. B., 'The "What" and "When" of self-initiated movement' (2012) 23(3) *Cerebral Cortex* 520.

Holroyd C. B. and Yeung N., 'Motivation of extended behaviours by anterior cingulate cortex' (2012) 16(2) *Trends in Cognitive Sciences* 122.

Horder J., 'A critique of the correspondence principle in criminal law' (1995) (Oct) *Criminal Law Review* 759.

Horder J., 'Reforming the auxiliary part of the criminal law' (2007) 10 *Archbold News* 6.

Horder J., 'Reshaping the subjective element in the provocation defence' (2005) 25(1) *Oxford Journal of Legal Studies* 123.

Horder J., 'Sobering up? The Law Commission on criminal intoxication' (1995) 58(4) *Modern Law Review* 534.

Horder J. and Fitz-Gibbon K., 'When sexual infidelity triggers murder: Examining the impact of homicide law reform on judicial attitudes in sentencing' (2015) 74(2) *Cambridge Law Journal* 307.

Hörnle T., 'Social expectations in the criminal law: The "reasonable person" in a comparative perspective' (2008) 11(1) *New Criminal Law Review: An International and Interdisciplinary Journal* 1.

Hosokawa T., Kennerley S. W., Sloan J. and Wallis J. D., 'Single-neuron mechanisms underlying cost-benefit analysis in frontal cortex' (2013) 33(44) *Journal of Neuroscience* 17385.

Houdayer E., Lee S.-J. and Hallett M., 'Cerebral preparation of spontaneous movements: An EEG study' (2020) 131(11) *Clinical Neurophysiology* 2561.

Houghton D. C., Capriotti M. R., Conelea C. A. and Woods D. W., 'Sensory phenomena in Tourette syndrome: Their role in symptom formation and treatment' (2014) 1(4) *Current Developmental Disorders Reports* 245.

Humphries M. D. and Gurney K. N., 'The role of intra-thalamic and thalamocortical circuits in action selection' (2002) 13(1) *Network Computation in Neural Systems* 131.

Humphries M. D., Steward R. D. and Gurney K. N., 'A physiologically plausible model of action selection and oscillatory activity in the basal ganglia' (2006) 26(50) *Journal of Neuroscience* 12921.

Ibbetson D., 'Recklessness restored' (2004) 63(1) *Cambridge Law Journal* 13.

Ireland J. L. and Higgins P., 'Behavioural stimulation and sensation-seeking among prisoners: Applications to substance dependency' (2013) 36(3-4) *International Journal of Law and Psychiatry* 229.

Izuma K., Akula S., Murayama K., Wu D.-A., Iacoboni M. and Adolphs R., 'A causal role for posterior medial frontal cortex in choice-induced preference change' (2015) 35(8) *Journal of Neuroscience* 3598.

Jackendoff R., 'How language helps us think' (1996) 4(1) *Pragmatics & Cognition* 1.

Jacoby L. L., Lindsay D. S. and Toth J., 'Unconscious influences revealed: Attention, awareness, and control' (1992) 47(6) *American Psychologist* 802.

Jahanshahi M. and Frith C. D., 'Willed action and its impairments' (1998) 15(6-8) *Cognitive Neuropsychology* 483.

Jahanshahi M., Jenkins H. I., Brown R. G., Marsden D. C., Passingham R. E. and Brooks D. J., 'Self-initiated versus externally triggered movements: I. An investigation using measurement of regional cerebral blood flow with PET and movement-related potentials in normal and Parkinson's disease subjects' (1995) 118(4) *Brain* 913.

Jeannerod M., 'The sense of agency and its disturbances in schizophrenia: A reappraisal' (2008) 192(3) *Experimental Brain Research* 527.

Jedlicka P., 'Revisiting the quantum brain hypothesis: Toward quantum (neuro)biology?' (2017) 10 *Frontiers in Molecular Neuroscience* 1.

Jefferson M., 'Recklessness: The objectivity of the Caldwell test' (1999) 63(1) *Journal of Criminal Law* 57.

Jo H.-G., Wittmann M., Hinterberger T. and Schmidt S., 'The readiness potential reflects intentional binding' (2014) 8 *Frontiers in Human Neuroscience* 421.

Johansson P., Hall L., Gulz A., Haake M. and Watanabe K., 'Choice blindness and trust in the virtual world' (2007) 107(60) *Technical Report of IEICE: HIP* 83.

Johansson P., Hall L. and Sikström S., 'From change blindness to choice blindness' (2008) 51(2) *Psychologia* 142.

Johansson P., Hall L., Sikström S. and Olsson A., 'Failure to detect mismatches between intention and outcome in a simple decision task' (2005) 310(5745) *Science* 116.

Johansson P., Hall L., Sikström S., Tärning B. and Lind A., 'How something can be said about telling more than we can know: on choice blindness and introspection' (2006) 15(4) *Consciousness and Cognition* 673.

Johansson P., Hall L., Tärning B., Sikström S. and Chater N., 'Choice blindness and preference change: You will like this paper better if you (believe you) chose to read it!' (2014) 27(3) *Journal of Behavioral Decision Making* 281.

Johnson R. E., Lin S.-H. and Lee H. W., 'Self-control as the fuel for effective self-regulation at work: Antecedents, consequences, and boundary conditions of employee self-control' (2018) 5 *Advances in Motivation Science* 87.

Johnstone C. L., 'An Aristotelian trilogy: Ethics, rhetoric, politics, and the search for moral truth' (1980) 13(1) *Philosophy & Rhetoric* 1.

Jones C. R., Fazio R. H. and Olson M. A., 'Implicit misattribution as a mechanism underlying evaluative conditioning' (2009) 96(5) *Journal of Personality and Social Psychology* 933.

Jonas K. J. and Sassenberg K., 'Knowing how to react: Automatic response priming from social categories' (2006) 90(5) *Journal of Personality and Social Psychology* 709.

Jones T. H., 'Insanity, automatism, and the burden of proof on the accused'' (1995) 111(Jul) *Law Quarterly Review* 475.

Josephson W. L., 'Television violence and children's aggression: Testing the priming, social script, and disinhibition predictions' (1987) 53(5) *Journal of Personality and Social Psychology* 882.

Kahneman D., 'A perspective on judgment and choice: Mapping bounded rationality' (2003) 58(9) *American Psychologist* 697.

Kahneman D. and Tversky A., 'Subjective probability: A judgment of representativeness' (1972) 3(3) *Cognitive Psychology* 430.

Kail R. and Salthouse T. A., 'Processing speed as a mental capacity' (1994) 86(2/3) *Acta Psychologica* 199.

Kalaska J. F., Scott S. H., Cisek P. and Sergio L. E., 'Cortical control of reaching movements' (1997) 7(6) *Neurobiology* 849.

Kalivas P. W. and Volkow N. D., 'The neural basis of addiction: A pathology of motivation and choice' (2005) 162(8) *American Journal of Psychiatry* 1403.

Kalven H., 'Insanity and the criminal law – A critique of *Durham v United States*' (1955) 22(2) *University of Chicago Law Review* 317.

Kantorowicz-Reznichenko E., 'Day-fines: Should the rich pay more?' (2015) 11(3) *Review of Law and Economics* 481.

Karp B. I., Porter S., Toro C. and Hallett M., 'Simple motor tics may be preceded by a premotor potential' (1996) 61(1) *Journal of Neurology, Neurosurgery and Psychiatry* 103.

Kaveny M. C., 'Inferring intention from foresight' (2004) 120(Jan) *Law Quarterly Review* 81.

Kawakami K., Phills C. E., Greenwald A. G., Simard D., Pontiero J., Brnjas A., Khan B., Mills J. and Dovidio J. F., 'In perfect harmony: Synchronizing the self to activated social categories' (2012) 102(3) *Journal of Personality and Social Psychology* 562.

Kearns G. and Mackay R. D., 'More fact(s) about the insanity defence' (1999) *Criminal Law Review* 714.

Kelly E. I., 'Criminal justice without retribution' (2009) 106(8) *Journal of Philosophy* 440.

Kennerley S. W., Dahmubed A. F., Lara A. H. and Wallis J. D., 'Neurons in the frontal lobe encode the value of multiple decision variables' (2009) 21(6) *Journal of Cognitive Neuroscience* 1162.

Kennett J. and Fine C., 'Will the real moral judgment please stand up? The implications of social intuitionist models of cognition for meta-ethics and moral psychology' (2009) 12(1) *Ethical Theory and Moral Practice* 77.

Kessler D. and Levitt S. D., 'Using sentence enhancements to distinguish between deterrence and incapacitation' (1999) 42(S1) *Journal of Law and Economics* 343.

Khalighinejad N., Costa S. D. and Haggard P., 'Endogenous action selection processes in dorsolateral prefrontal cortex contribute to sense of agency: A meta-analysis of tDCS studies of "intentional binding"' (2016) 9(3) *Brain Stimulation* 372.

Khalighinejad N. and Haggard P., 'Modulating human sense of agency with non-invasive brain stimulation' (2015) 69 *Cortex* 93.

Kidd C., Palmeri H. and Aslin R. N., 'Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability' (2013) 126(1) *Cognition* 109.

Kimel D., 'Inadvertent recklessness in criminal law' (2004) 120(Oct) *Law Quarterly Review* 548.

Klaes C., Westendorff S., Chakrabarti S. and Gail A., 'Choosing goals, not rules: Deciding among rule-based actions plans' (2011) 70(3) *Neuron* 536.

Kleinlogel E. P., Dietz J. and Antonakis J., 'Lucky, competent, or just a cheat? Interactive effects of honesty-humility and moral cues on cheating behavior' (2017) 44(2) *Personality and Social Psychology Bulletin* 158.

Klineberg S. L., 'Future time perspective and the preference for delayed reward' (1968) 8(3) *Journal of Personality and Social Psychology* 253.

Knapman L., 'Murder – dangerous act – foresight of death or serious bodily harm' (1986) (Nov) *Criminal Law Review* 742.

Knapman L., 'Permitting use of premises as unlicensed sex establishment' (1986) (Oct) *Criminal Law Review* 693.

Koi P., Uusitalo S. and Tuominen J., 'Self-control in responsibility enhancement and criminal rehabilitation' (2018) 12(2) *Criminal Law and Philosophy* 227.

Koob G. F. and Le Moal M., 'Drug abuse: Hedonic homeostatic dysregulation' (1997) 278(5335) *Science* 52.

Koriat A. and Feuerstein N., 'The recovery of incidentally acquired information' (1976) 40(6) *Acta Psychologica* 463.

Kornhuber H. H. and Deecke L., 'Hirnpotentialänderungen bei willkürbewegungen und passiven bewegungen des menschen: Bereitschaftspotential und reafferente potentiale' (1965) 284(1) *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere* 1.

Kornhuber H. H. and Deecke L., 'Readiness for movement – The bereitschaftspotential story' (1990) 33(4) *Current Contents Life Sciences* 14.

Kotecha B., 'Necessity as a defence to murder: an Anglo-Canadian perspective' (2014) 78(4) *Journal of Criminal Law* 341.

Kouneiher F., Charron S. and Koechlin E., 'Motivation and cognitive control in the human prefrontal cortex' (2009) 12(7) *Nature Neuroscience* 939.

Kovandzic T., Sloan J. J. and Vieraitis L. M., 'Unintended consequences of politically popular sentencing policy: The homicide promoting effects of "three-strikes" in US cities 1980 – 1999' (2006) 1(3) *Criminology and Public Policy* 399.

Kraus B. J., Robinson II R. J., White J. A., Eichenbaum H. and Hasselmo M. E., 'Hippocampal "time cells": Time versus path integration' (2013) 78(6) *Neuron* 1090.

Krebs B., 'Oblique intent, foresight and authorisation' (2018) 7(2) *UCL Journal of Law and Jurisprudence* 1.

Krieghoff V., Brass M., Prinz W. and Waszak F., 'Dissociating what and when of intentional actions' (2009) 3 *Frontiers in Human Neuroscience* 1.

Krieghoff V., Waszak F., Prinz W. and Brass M., 'Neural and behavioral correlates of intentional actions' (2011) 49(5) *Neuropsychologia* 767.

Kuhn D., 'Thinking as argument' (1992) 62(2) *Harvard Educational Review* 155.

Kühn S. and Brass M., 'Retrospective construction of the judgment of free choice' (2009) 18(1) *Consciousness and Cognition* 12.

Kühn S., Haggard P. and Brass M., 'Intentional inhibition: How the "veto-area" exerts control' (2009) 30(9) *Human Brain Mapping* 2834.

Kuhn L. J., Willoughby M. T., Vernon-Feagans L., Blair C. B. and Family Life Project Key Investigators, 'The contributions of children's time-specific and longitudinal expressive language skills on developmental trajectories of executive function' (2016) 148 *Journal of Experimental Child Psychology* 20.

Kunda Z., 'The case for motivated reasoning' (1990) 108(3) *Psychological Bulletin* 480.

Krugwasser A. R., Stern Y., Faivre N., Harel E. V. and Salomon R., 'Impair sense of agency and associated confidence in psychosis' (2022) 8(32) *Schizophrenia* 1.

Lambert T. B., 'Theft, homicide and crime in late Anglo-Saxon law' (2012) 214(1) *Past & Present* 3.

Lacey N., 'A clear concept of intention: Elusive or illusory?' (1993) 56(5) *Modern Law Review* 621.

Lacey N., 'Responsibility without consciousness' (2015) 36(2) *Oxford Journal of Legal Studies* 219.

Laird K., 'Dishonesty: *R v Barton*; *R v Booth;* Court of Appeal: 29 April 2020' (2020) 11 *Criminal Law Review* 1065.

Laird K., 'Homicide: *R v Golds (Mark Richard)* Supreme Court' (2017) 4 *Criminal Law Review* 316.

Laird K., 'Terrorism: *R v Lane (Sally)* Supreme Court: Lady Hale PSC, Lord Burnett CJ, Lords Hughes, Hodge and Mance: 11 July 2018' (2019) 2 *Criminal Law Review* 178.

Lamm B., Keller H., Teiser J., Gudi H., Yovsi R. D., Freitag C., Poloczek S., Fassbender I., Suhrka J., Teubert M., Vöhringer I., Knoopf M., Schwarzer G. and Lohaus A., 'Waiting for the second treat: Developing culture-specific models of self-regulation' (2018) 89(3) *Child Development* e261.

Landenberger N. A. and Lipsey M. W., 'The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment' (2005) 1(4) *Journal of Experimental Criminology* 451.

Lansink C. S., Jackson J. C., Lankelma J. V., Ito R., Robbins T. W., Everitt B. J. and Pennartz C. M. A., 'Reward cues in space: Commonalities and differences in neural coding by hippocampal and ventral striatal ensembles' (2012) 32(36) *Journal of Neuroscience* 12444.

Lappi-Seppälä T., 'The fall of the Finnish prison population' (2000) 1(1) *Journal of Scandanavian Studies in Criminology and Crime Prevention* 27.

Latimer J., Dowden C. and Muise D., 'The effectiveness of restorative justice practices: A meta-analysis' (2005) 85(2) *The Prison Journal* 127.

Laughlin P. R. and Ellis A. L., 'Demonstrability and social combination processes on mathematical intellective tasks' (1986) 22(3) *Journal of Experimental Social Psychology* 177.

Lau H. C. and Passingham R. E., 'Unconscious activation of the cognitive control system in the human prefrontal cortex' (2007) 27(21) *Journal of Neuroscience* 5805.

Lau H. C., Rogers R. D., Haggard P. and Passingham R. E., 'Attention to intention' (2004a) 303(5661) *Science* 1208.

Lau H. C., Rogers R. D. and Passingham R. E., 'Dissociating response selection and conflict in the medial frontal surface' (2006) 29(2) *NeuroImage* 446.

Lau H. C., Rogers R. D., Ramnani N. and Passingham R. E., 'Willed action and attention to the selection of action' (2004b) 21(4) *NeuroImage* 1407.

Ledgerwood A. and Chaiken S., 'Priming us and them: Automatic assimilation and contrast in group attitudes' (2007) 93(6) *Journal of Personality and Social Psychology* 940.

LeDoux J. E., Wilson D. H. and Gazzaniga M. S., 'A divided mind: Observations on the conscious properties of the separated hemispheres' (1977) 2(5) *Annals of Neurology* 417.

Lee B. M. and Kemmelmeier M., 'How reliable are the effects of self-control training?: A re-examination using self-report and physical measures' (2017) 12(6) *PLoS ONE* e0178814.

Lee D. S. and McCrary J., 'The deterrence effect of prison: Dynamic theory and evidence' (2017) 38 *Advances in Econometrics* 73.

Leggett Z., 'The new test for dishonesty in criminal law – Lessons from the Courts of Equity?' (2020) 84(1) *Journal of Criminal Law* 37.

Lens J. W., 'Justice Holmes's bad man and the depleted purposes of punitive damages' (2013) 101(4) *Kentucky Law Journal* 789.

Lerner J. S., Goldberg J. H. and Tetlock P. E., 'Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility' (1998) 24(6) *Personality and Social Psychology Bulletin* 563.

Leutgeb S., Leutgeb J. K., Barnes C. A., Moser E. I., McNaughton B. L. and Moser M.-B., 'Independent codes for spatial and episodic memory in hippocampal neuronal ensembles' (2005) 309(5734) *Science* 619.

Leverick F., 'Is English self-defence law incompatible with Article 2 of the ECHR?' (2002) (May) *Criminal Law Review* 347.

Leverick F., 'Unreasonable mistake in self-defence: *Liester v HM Advocate*' (2009) 13(1) *Edinburgh Law Review* 100.

Liberman N. and Förster J., 'Expression after suppression: A motivational explanation of postsupressional rebound' (2000) 79(2) *Journal of Personality and Social Psychology* 190.

Libet B., 'Brain stimulation in the study of neuronal functions for conscious sensory experience' (1982) 1(4) *Human Neurobiology* 235.

Libet B., 'Cerebral physiology of conscious experience: Experimental studies in human subjects' in Osaka N. (ed.), *Neural Basis of Consciousness* (John Benjamins Publishing 2003).

Libet B., 'Do we have free will?' (1999) 6(8-9) *Journal of Consciousness Studies* 47.

Libet B., 'Unconscious cerebral initiative and the role of conscious will in voluntary action' (1985) 8(4) *Behavioral and Brain Sciences* 529.

Libet B., Alberts W. W., Wright Jr E. W., Delattre L. D., Levin G. and Feinstein B., 'Production of threshold levels of conscious sensation by electrical stimulation of human somatosensory cortex' (1964) 27(4) *Journal of Neurophysiology* 546.

Libet B., Alberts W. W., Wright Jr E. W. and Feinstein B., 'Responses of human somatosensory cortex stimuli below threshold for conscious sensation' (1967) 158(3808) *Science* 1597.

Libet B., Wright Jr E. W., Feinstein B. and Pearl D. K., 'Retroactive enhancement of a skin sensation by a delayed cortical stimulus in man: Evidence for delay of a conscious sensory experience' (1992) 1(3) *Consciousness and Cognition* 367.

Libet B., Wright Jr E. W., Feinstein B. and Pearl D. K., 'Subjective referral of the timing for a conscious sensory experience: A functional role for the somatosensory specific projection system in man' (1979) 102(1) *Brain: A Journal of Neurology* 193.

Libet B., Wright Jr E. W. and Gleason C. A., 'Readiness-potentials preceding unrestricted "spontaneous" vs pre-planned voluntary acts' (1982) 54(3) *Electroencephalography and Clinical Neurophysiology* 322.

Libet B., Wright Jr E. W. and Gleason C. A., 'Preparation- or intention-to-act, in relation to pre-event potentials recorded at the vertex' (1983a) 56(4) *Electroencephalography and Clinical Neurophysiology* 367.

Libet B., Wright Jr E. W., Gleason C. A. and Pearl D. K., 'Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act' (1983b) 106(3) *Brain: A Journal of Neurology* 623.

Liljeholm M. and O'Doherty J. P., 'Contributions of the striatum to learning, motivation, and performance: an associative account' (2012) 16(9) *Trends in Cognitive Sciences* 467.

Lim J., Ho P. M. and Mullette-Gillman O.'D. A., 'Modulation of incentivized dishonesty by disgust facial expressions' (2015) 9 *Frontiers in Neuroscience* 250.

Lio J., '*Cunningham* recklessness: The quintessence of the historic English criminal law?' (2018) 6 *North East Law Review* 71.

Lipsey M. W. and Cullen F. T., 'Correctional rehabilitation: A review of systematic reviews' (2007) 3(1) *Annual Review of Law and Social Science* 297.

Loeffler C. E. and Nagin D. S., 'The impact of incarceration on recidivism' (2022) 5(1) *Annual Review of Criminology* 133.

Loersch C., Aarts H., Payne B. K. and Jefferis V. E., 'The influence of social groups on goal contagion' (2008) 44(6) *Journal of Experimental Social Psychology* 1555.

Loersch C., Durso G. R. O. and Petty R. E., 'Vicissitudes of desire: A matching mechanism for subliminal persuasion' (2013) 4(5) *Social Psychological and Personality Science* 624.

Loersch C. and Payne B. K., 'On mental contamination: The Role of (mis)attribution in behavior priming' (2012) 30(2) *Social Cognition* 241.

Loersch C. and Payne B. K., 'Situated inferences and the what, who, and where of priming' (2014) 32(Supp) *Social Cognition* 137.

Loersch C. and Payne B. K., 'The situated inference model: An integrative account of the effects of primes on perception, behavior, and motivation' (2011) 6(3) *Perspectives on Psychological Science* 234.

Logan G. D. and Cowan W. B., 'On the ability to inhibit thought and action: A theory of an act of control' (1984) 91(3) *Psychological Review* 295.

Lombardi W. J., Higgins E. T. and Bargh J. A., 'The role of consciousness in priming effects on categorization: Assimilation versus contrast as a function of awareness of the priming task' (1987) 13(3) *Personality and Social Psychology Bulletin* 411.

Loughnan A., '"Manifest madness": Towards a new understanding of the insanity defence' (2007) 70(3) *Modern Law Review* 379.

Lu L. H., Barrett A. M., Schwartz R. L., Cibula J. E., Gilmore R. L., Uthman B. M. and Heilman K. M., 'Anosognosia and confabulation during the Wada test' (1997) 49(5) *Neurology* 1316.

Ma F., Chen B., Xu F., Lee K. and Heyman G. D., 'Generalized trust predicts young children's willingness to delay gratification' (2018) 169 *Journal of Experimental Child Psychology* 118.

MacDonald P. A. and Paus T., 'The role of parietal cortex in awareness of self-generated movements: A transcranial magnetic stimulation study' (2003) 13(9) *Cerebral Cortex* 962.

Maciejovsky B. and Budescu D. V., 'Collective induction without cooperation? Learning and knowledge transfer in cooperative groups and competitive auctions' (2007) 92(5) *Journal of Personality and Social Psychology* 854.

Mackay R. D., 'An anatomy of automatism' (2015) 55(3) *Medicine, Science and the Law* 150.

Mackay R. D., '"Nature", "quality" and *mens rea* – Some observations on "defect of reason" and the first limb of the M'Naghten rules' (2020) 7 *Criminal Law Review* 588.

Mackay R. D., 'R v B: Rape – Consent – Defendant suffering from mental illness at time of offence Court of Appeal' (2014) 4 *Criminal Law Review* 312.

Mackay R. D., 'The abnormality of mind factor in diminished responsibility' (1999) (Feb) *Criminal Law Review* 117.

Mackay R. D. and Mitchell B. J., 'Sleepwalking, automatism and insanity' (2006) (Oct) *Criminal Law Review* 901.

Mackay R. D. and Reuber M., 'Epilepsy and the defence of insanity: Time for change?' (2007) (Oct) *Criminal Law Review* 782.

Magill K., 'The idea of justification for punishment' (1998) 1(1) *Critical Review of International Social and Political Philosophy* 86.

Marcel A. J., 'Conscious and unconscious perception: Experiments on visual masking and word recognition' (1983) 15(2) *Cognitive Psychology* 197.

Marien H., Custers R., Hassin R. R. and Aarts H., 'Unconscious goal activation and the hijacking of the executive function' (2012) 103(3) *Journal of Personality and Social Psychology* 399.

Markus H. and Kunda Z., 'Stability and malleability of the self-concept' (1986) 51(4) *Journal of Personality and Social Psychology* 858.

Markus H. and Wurf E., 'The dynamic self-concept: A social psychological perspective' (1987) 38(1) *Annual Review of Psychology* 299.

Marks V., 'Hypoglycaemia and Automatism' (2015) 55(3) *Medicine, Science and the Law* 186.

Marneweck M. and Flamand V. H., 'Elucidating the neural circuitry underlying planning of internally-guided voluntary action' (2016) 116(6) *Journal of Neurophysiology* 2469.

Marsh K. and Fox C., 'The benefit and cost of prison in the UK. The results of a model of lifetime re-offending' (2008) 4(4) *Journal of Experimental Criminology* 403.

Martin J.-R., 'Experiences of activity and causality in schizophrenia: When predictive deficits lead to a retrospective over-binding' (2013) 22(4) *Consciousness and Cognition* 1361.

Marvell T. B. and Moody C. E., 'The lethal effects of three-strikes laws' (2001) 30(1) *Journal of Legal Studies* 89.

Mascaro O. and Sperber D., 'The moral, epistemic, and mindreading components of children's vigilance towards deception' (2009) 112(3) *Cognition* 367.

Mazar N., Amir O. and Ariely D., 'The dishonesty of honest people: A theory of self-concept maintenance' (2008) 45(6) *Journal of Marketing Research* 633.

McAuley F., 'The grammar of mistake in criminal law' (1996) 31 *Irish Jurist* 56.

McClanahan W. P. and van der Linden S., 'An uncalculated risk: Ego-depletion reduces the influence of perceived risk but not state affect on criminal choice' (2020) *Psychology, Crime & Law* 1.

McEwan B. S., 'Protective and damaging effects of stress mediators' (1998) 338(3) *New England Journal of Medicine* 171.

McEwan J., '"I thought she consented": Defeat of the rape shield or the defence that shall not run?' (2006) (Nov) *Criminal Law Review* 969.

McFarland C. P. and Glisky E. L., 'Frontal lobe involvement in a task of time-based prospective memory' (2009) 47(7) *Neuropsychologia* 1660.

McKenna M., 'A hard-line reply to Pereboom's four-case manipulation argument' (2008a) 77(1) *Philosophy and Phenomenological Research* 142.

McKenna M., 'Saying good-bye to the Direct Argument the right way' (2008b) 117(3) *The Philosophical Review* 349.

McPeek R. M. and Keller E. L., 'Superior colliculus activity related to concurrent processing of saccade goals in a visual search task' (2002) 87(4) *Journal of Neurophysiology* 1805.

McFadden J., 'The conscious electromagnetic information (cemi) field theory' (2002) 9(8) *Journal of Consciousness Studies* 45.

Mead N. L., Baumeister R. F., Gino F., Schweitzer M. E. and Ariely D., 'Too tired to tell the truth: Self-control resource depletion and dishonesty' (2009) 45(3) *Journal of Experimental Social Psychology* 594.

Mears D. P., Cochran J. C., Bales W. D. and Bhati A. S., 'Recidivism and time served in prison' (2016) 106(1) *Journal of Criminal Law and Criminology* 81.

Mele A., 'Unconscious decisions and free will' (2013) 26(6) *Philosophical Psychology* 777.

Mercier H. and Sperber D., 'Why do humans reason? Arguments for an argumentative theory' (2011) 34(2) *Behavioral and Brain Sciences* 57.

Meyer D. E. and Schvaneveldt R. W., 'Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations' (1971) 90(2) *Journal of Experimental Psychology* 227.

Michaelson L. E. and Munakata Y., 'Same data set, different conditions: Preschool delay of gratification predicts later behavioral outcomes in a preregistered study' (2020) *Psychological Science* 1.

Michely J., Volz L. J., Barbe M. T., Hoffstaedter F., Viswanathan S., Timmermann L., Eickhoff S. B., Fink G. R. and Grefkes C., 'Dopaminergic modulation of motor network dynamics in Parkinson's disease' (2015) 138(3) *Brain* 664.

Mikhail J. M., 'Law, science, and morality: A review of Richard Posner's "The problematics of moral and legal theory"' (2002) 54(5) *Stanford Law Review* 1057.

Mikhail J. M., 'Universal moral grammar: Theory, evidence and the future' (2007) 11(4) *Trends in Cognitive Sciences* 143.

Mikhail J. M., Sorrentino C. M. and Spelke E. S., 'Toward a universal moral grammar' in Gernsbacher M. A. and Derry S. J. (eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (Lawrence Erlbaum Associates 1998).

Miles T. J. and Ludwig J., 'The silence of the Lambdas: Deterring incapacitation research' (2007) 23(4) *Journal of Quantitative Criminology* 287.

Mills C., 'Politics and manipulation' (1995) 21(1) *Social Theory and Practice* 97.

Mirabella G., 'Endogenous inhibition and the neural basis of "free won't"' (2007) 27(51) *Journal of Neuroscience* 13919.

Mischel W., 'Delay of gratification, need for achievement, and acquiescence in another culture' (1961) 62(3) *Journal of Abnormal and Social Psychology* 543.

Mischel W., 'Preference for delayed reinforcement: An experimental study of a cultural observation' (1958) 56(1) *Journal of Abnormal and Social Psychology* 57.

Mischel W., 'Preference for delayed reinforcement and social responsibility' (1961) 62(1) *Journal of Abnormal and Social Psychology* 1.

Mischel W., 'Processes in delay of gratification' (1974) 7 *Advances in Experimental Social Psychology* 249.

Mischel W. and Ebbesen E. B., 'Attention in delay of gratification' (1970) 16(2) *Journal of Personality and Social Psychology* 329.

Mischel W., Ebbesen E. B. and Zeiss A. R., 'Cognitive and attentional mechanisms in delay of gratification' (1972) 21(2) *Journal of Personality and Social Psychology* 204.

Mischel W., Shoda Y. and Peake P. K., 'The nature of adolescent competencies predicted by preschool delay of gratification' (1988) 54(4) *Journal of Personality and Social Psychology* 687.

Mitchell B., 'Culpably indifferent murder' (1996) 25(1) *Anglo-American Law Review* 64.

Mitchell B., 'Recklessness could still be a state of mind' (1988) 52(3) *Journal of Criminal Law* 300.

Mitchell B. and Mackay R. D., 'The gold standard of substantial impairment' (2015) 4 *Archbold Review* 7.

Mitchell C. N., 'The intoxicated offender – Refuting the legal and medical myths' (1988) 11(1) *International Journal of Law and Psychiatry* 77.

Mitchell M. S., Baer M. D., Ambrose M. L., Folger R. and Palmer N. F., 'Cheating under pressure: A self-protection model of workplace cheating behavior' (2018) 103(1) *Journal of Applied Psychology* 54.

Miyake A. and Friedman N. P., 'The nature and organization of individual differences in executive functions: Four general conclusions' (2012) 21(1) *Current Directions in Psychological Science* 8.

Molden D. C., 'Understanding priming effects in social psychology: What is "social priming" and how does it occur?' (2014) 32(Supp) *Social Cognition* 1.

Momennejad I. and Haynes J.-D., 'Human anterior prefrontal cortex encodes the "what" and "when" of future intentions' (2012) 61(1) *NeuroImage* 139.

Monteith M. J., Ashburn-Nardo L., Voils C. I. and Czopp A. M., 'Putting the brakes on prejudice: On the development and operation of cues for control' (2002) 83(5) *Journal of Personality and Social Psychology* 1029.

Moore J. W. and Fletcher P. C., 'Sense of agency in health and disease: A review of cue integration approaches' (2012) 21(1) *Consciousness and Cognition* 59.

Moore J. W. and Haggard P., 'Awareness of action: Inference and prediction' (2008) 17(1) *Consciousness and Cognition* 136.

Moore J. W. and Obhi S. S., 'Intentional binding and the sense of agency: A review' (2012) 21(1) *Consciousness and Cognition* 546.

Moore M. S., 'A tale of two theories' (2009) 28(1) *Criminal Justice Ethics* 27.

Moore M. S., 'Choice, character, and excuse' (1990) 7(2) *Social Philosophy and Policy* 29.

Moore M. S., 'Four reflections on law and morality' (2007) 48(5) *William & Mary Law Review* 1523.

Moore M. S., 'Justifying retributivism' (1993) 27(1-2) *Israel Law Review* 15.

Moore M. S., 'The various relations between law and morality in contemporary philosophy' (2012) 25(4) *Ratio Juris* 435.

Moore M. S. and Hurd H. M., 'Punishing the awkward, the stupid, the weak, and the selfish: The culpability of negligence' (2011) 5(2) *Criminal Law and Philosophy* 147.

Moretti R. and Signori R., 'Neural correlates for apathy: Frontal-prefrontal and parietal cortical- subcortical circuits' (2016) 8 *Frontiers in Aging Neuroscience* 1.

Moreno-Bote R., Rinzel J. and Rubin N., 'Noise-induced alternations in an attractor network model of perceptual bistability' (2007) 98(3) *Journal of Neurophysiology* 1125.

Morris N., 'Somnambulistic homicide' (1951) 5(1) *Res Judicatae* 29.

Morrison K. R., Johnson C. S. and Wheeler S. C., 'Not all selves feel the same uncertainty: Assimilation to primes among individualists and collectivists' (2012) 3(1) *Social Psychological and Personality Science* 118.

Morrison S. F., Nakamura K. and Madden C. J., 'Central control of thermogenesis in mammals' (2008) 93(7) *Experimental Physiology* 773.

Morse S. J., 'Causation, compulsion, and involuntariness' (1994) 22(2) *Bulletin of the American Academy of Psychiatry and the Law* 159.

Morse S. J., 'Determinism and the death of folk psychology: Two challenges to responsibility from neuroscience' (2008) 9(1) *Minnesota Journal of Law, Science & Technology* 1.

Morse S. J., 'The non-problem of free will in forensic psychiatry and psychology' (2007) 25(2) *Behavioral Sciences & the Law* 203.

Moshman D. and Geil M., 'Collaborative reasoning: Evidence for collective rationality' (1998) 4(3) *Thinking & Reasoning* 231.

Mueller V. A., Brass M., Waszak F. and Prinz W., 'The role of the preSMA and the rostral cingulate zone in internally selected actions' (2007) 37(4) *NeuroImage* 1354.

Muller R. U. and Kubie J. L., 'The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells' (1987) 7(7) *Journal of Neuroscience* 1951.

Munby J., 'Law, morality and religion in the family courts' (2014) 16(2) *Ecclesiastical Law Journal* 131.

Muraven M. and Baumeister R. F., 'Self-regulation and depletion of limited resources: Does self-control resemble a muscle?' (2000) 126(2) *Psychological Bulletin* 247.

Muraven M., Baumeister R. F. and Tice D. M., 'Longitudinal improvement of self-regulation through practice: Building self-control strength through repeated exercise' (1999) 139(4) *Journal of Social Psychology* 446.

Muraven M., Collins R. L. and Nienhaus K., 'Self-control and alcohol restraint: An initial application of the self-control strength model' (2002) 16(2) *Psychology of Addictive Behaviors* 113.

Muraven M., Pogarsky G. and Shmueli D., 'Self-control depletion and the general theory of crime' (2006) 22(3) *Journal of Quantitative Criminology* 263.

Muraven M., Tice D. M. and Baumeister R. F., 'Self-control as limited resource: Regulatory depletion patterns' (1998) 74(3) *Journal of Personality and Social Psychology* 774.

Murray M. J. and Dudrick D. F., 'Are coerced acts free?' (1995) 32(2) *American Philosophical Quarterly* 109.

Mussweiler T. and Neumann R., 'Sources of mental contamination: Comparing the effects of self-generated versus externally provided primes' (2000) 36(2) *Journal of Experimental Social Psychology* 194.

Naccache L. and Dehaene S., 'The priming method: Imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes' (2001) 11(10) *Cerebral Cortex* 966.

Nachev P., Rees G., Parton A., Kennard C. and Husain M., 'Volition and conflict in human medial frontal cortex' (2005) 15(2) *Current Biology* 122.

Nachev P., Wydell H., O'Neill K., Husain M. and Kennard C., 'The role of the pre-supplementary motor area in the control of action' (2007) 36(3) *NeuroImage* T155.

Nagin D. S., Cullen F. T. and Jonson C. L., 'Imprisonment and reoffending' (2009) 38(1) *Crime and Justice* 115.

Nann M., Cohen L. G., Deecke L. and Soekadar S. R., 'To jump or not to jump – The bereitschaftspotential required to jump into 192-meter abyss' (2019) 9(1) *Scientific Reports* 2243.

Neely J. H., 'Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention' (1977) 106(3) *Journal of Experimental Psychology* 226.

Nelson S. A., 'Factors influencing young children's use of motives and outcomes as moral criteria' (1980) 51(3) *Child Development* 823.

Newell B. R. and Shanks D. R., 'Prime numbers: Anchoring and its implications for theories of behavior priming' (2014) 32(Supp) *Social Cognition* 88.

Nguyen V. T., Breakspear M. and Cunnington R., 'Reciprocal interactions of the SMA and cingulate cortex sustain premovement activity for voluntary actions' (2014) 34(49) *Journal of Neuroscience* 16397.

Nichols A. D., Lang M., Kavanagh C., Kundt R., Yamada J., Ariely D. and Mitkidis P., 'Replicating and extending the effects of auditory religious cues on dishonest behavior' (2020) 15(8) *PLoS ONE* e0237007.

Nisbett R. E. and Wilson T. D., 'Telling more than we can know: Verbal reports on mental processes' (1977) 84(3) *Psychological Review* 231.

Niv Y. and Shoenbaum G., 'Dialogues on prediction errors' (2008) 12(7) *Trends in Cognitive Sciences* 265.

Oaten M. and Cheng K., 'Improved self-control: The benefits of a regular program of academic study' (2006a) 28(1) *Basic and Applied Social Psychology* 1.

Oaten M. and Cheng K., 'Improvements in self-control from financial monitoring' (2007) 28(4) *Journal of Economic Psychology* 487.

Oaten M. and Cheng K., 'Longitudinal gains in self-regulation from regular physical exercise' (2006b) 11(4) *British Journal of Health Psychology* 717.

Oettingen G., Grant H., Smith P. K., Skinner M. and Gollwitzer P. M., 'Nonconscious goal pursuit: Acting in an explanatory vacuum' (2006) 42(5) *Journal of Experimental Social Psychology* 668.

O'Keefe J. and Dostrovsky J., 'The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat' (1971) 34(1) *Brain Research* 171.

Oliphant S. N., 'Estimating the effect of death penalty moratoriums on homicide rates using the synthetic control method' (2022) 0(0) [online] *Criminology and Public Policy* 1.

Omata K., Ito S., Takata Y. and Ouchi Y., 'Similar neural correlates of planning and execution to inhibit continuing actions' (2018) 12 *Frontiers in Neuroscience* 1.

Ong H. H., Mullette-Gillman O.'D. A., Kwok K. and Lim J., 'Moral judgment modulation by disgust is bi-directionally moderated by individual sensitivity' (2014) 5 *Frontiers in Psychology* 194.

Oosterhof N. N., Tipper S. P. and Downing P. E., 'Visuo-motor imagery of specific manual actions: a multi-variate pattern analysis fMRI study' (2012) 63(1) *NeuroImage* 262.

Ordóñez L. D. and Welsh D. T., 'Immoral goals: How goal setting may lead to unethical behavior' (2015) 6 *Current Opinion in Psychology* 93.

O'Reilly R., 'The what and how of prefrontal cortical organization' (2010) 3(8) *Trends in Neurosciences* 355.

Ormerod D., 'Necessity of circumstance' (2006) (Feb) *Criminal Law Review* 148.

Ormerod D., 'Proceeds of crime: Assisting another to retain benefit of criminal conduct knowing or suspecting other person to be engaged in criminal conduct' (2007) (Jan) *Criminal Law Review* 77.

Ormerod D., 'Trial: Direction to jury – Reasonable person – Reasonable conduct – Defendant suffering from paranoid schizophrenia' (2001) (Oct) *Criminal Law Review* 845.

Ormerod D. and Laird K., 'The future of dishonesty – Some practical considerations' (2020) 6 *Archbold Review* 8.

Owusu G. M. Y., Bekoe R. A., Koomson T. A. A. and Simpson S. N. Y., 'Temptation and the propensity to engage in unethical behaviour' (2018) 35(1) *International Journal of Ethics and Systems* 43.

Padoa-Schioppa C., 'Range-adapting representation of economic value in the orbitofrontal cortex' (2009) 29(44) *Journal of Neuroscience* 14004.

Padoa-Schioppa C. and Assad J. A., 'Neurons in the orbitofrontal cortex encode economic value' (2006) 441(7090) *Nature* 223.

Paik H. and Comstock G., 'The effects of television violence on antisocial behavior: A meta-analysis' (1994) 21(4) *Communication Research* 516.

Panee C. D. and Ballard M. E., 'High versus low aggressive priming during videogame training: Effects on violent action during game play, hostility, heart rate, and blood pressure' (2002) 32(12) *Journal of Applied Social Psychology* 2458.

Papachristos A. V., Meares T. L. and Fagan J., 'Why do criminals obey the law? The influence of legitimacy and social networks on active gun offenders' (2012) 102(2) *Journal of Criminal Law and Criminology* 397.

Pareés I., Brown H., Nuruki A., Adams R. A., Davare M., Bhatia K. P., Friston K. and Edwards M. J., 'Loss of sensory attenuation in patients with functional (psychogenic) movement disorders' (2014) 137(11) *Brain* 2916.

Parhar K. K, Wormith S. J., Derkzen D. M. and Beauregard A. M., 'Offender coercion in treatment: A meta-analysis of effectiveness' (2008) 35(9) *Criminal Justice and Behavior* 1109.

Parkinson J. and Haggard P., 'Choosing to stop: Responses evoked by externally triggered and internally generated inhibition identify a neural mechanism of will' (2015) 27(10) *Journal of Cognitive Neuroscience* 1948.

Parkinson J. and Haggard P., 'Subliminal priming of intentional inhibition' (2014) 130(2) *Cognition* 255.

Parks-Stamm E. J., Oettingen G. and Gollwitzer P. M., 'Making sense of one's actions in an explanatory vacuum: The interpretation of nonconscious goal striving' (2010) 46(3) *Journal of Experimental Social Psychology* 531.

Parsons S., 'The loss of control defence – Fit for purpose?' (2015) 79(2) *Journal of Criminal Law* 94.

Pastor-Bernier A. and Cisek P., 'Neural correlates of biased competition in premotor cortex' (2011) 31(19) *Journal of Neuroscience* 7083.

Pastor-Barnier A., Tremblay E. and Cisek P., 'Dorsal premotor cortex is involved in switching motor plans' (2012) 5(5) *Frontiers in Neuroengineering* 1.

Patterson D., 'Rethinking duress' (2016) 7(3) *Jurisprudence* 672.

Patient I. H. E., 'Mistake of law – A mistake?' (1987) 51(3) *Journal of Criminal Law* 326.

Paul G. M. and Lange K. W., 'Epilepsy and criminal law' (1992) 32(2) *Medicine, Science and the Law* 160.

Pazzaglia M. and Galli G., 'Loss of agency in apraxia' (2014) 8 *Frontiers in Human Neuroscience* 751.

Pearson F. S., Lipton D. S., Cleland C. M. and Yee D. S., 'The effects of behavioral/cognitive-behavioral programs on recidivism' (2002) 48(3) *Crime and Delinquency* 476.

Pedain A., 'Intention and the terrorist example' (2003) (Sep) *Criminal Law Review* 579.

Peña J., Hancock J. T. and Merola N. A., 'The priming effects of avatars in virtual settings' (2009) 36(6) *Communication Research* 838.

Pedersen W. C., Vasquez E. A., Bartholow B. D., Grosvenor M. and Truong A., 'Are you insulting me? Exposure to alcohol primes increases aggression following ambiguous provocation' (2014) 40(8) *Personality and Social Psychology Bulletin* 1037.

Peña J., McGlone M. S. and Sanchez J., 'The cowl makes the monk: How avatar appearance and role labels affect cognition in virtual worlds' (2012) 5(3) *Journal for Virtual Worlds Research* 1.

Pereboom D., 'Determinism al dente' (1995) 29(1) *Noûs* 21.

Pettigrew T. F., 'Intergroup contact theory' (1998) 49(1) *Annual Review of Psychology* 65.

Pettigrew T. F. and Tropp L. R., 'A meta-analytic test of intergroup contact theory' (2006) 90(5) *Journal of Personality and Social Psychology* 751.

Petty R. E. and Cacioppo J. T., 'Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses' (1979) 37(10) *Journal of Personality and Social Psychology* 1915.

Pezzulo G., van der Meer M. A. A., Lansink C. S. and Pennartz C. M. A., 'Internally generated sequences in learning and executing goal-directed behavior' (2014) 18(12) *Trends in Cognitive Sciences* 647.

Pfeiffer B. E. and Foster D. J., 'Hippocampal place-cell sequences depict future paths to remembered goals' (2013) 497(7447) *Nature* 74.

Pigott M., 'Intention – murder – model direction laid down in Moloney unsafe and misleading' (1986) (Jun) *Criminal Law Review* 400.

Pigott M., 'Murder – intention' (1986) (Mar) *Criminal Law Review* 180.

Piquero A. R., MacDonald J., Dobrin A., Daigle L. E. and Cullen F. T., 'Self-control, violent offending, and homicide victimization: Assessing the general theory of crime' (2005) 21(1) *Journal of Quantitative Criminology* 55.

Pizarro D. A. and Bloom P., 'The intelligence of the moral intuition: Comment on Haidt' (2003) 110(1) *Psychological Review* 193.

Platt M. L. and Glimcher P. W., 'Neural correlates of decision variables in parietal cortex' (1999) 400(6741) *Nature* 233.

Pocheptsova A., Amir O., Dhar R. and Baumeister R. F., 'Deciding without resources: Resource depletion and choice in context' (2009) 46(3) *Journal of Marketing Research* 344.

Polinsky A. M. and Shavell S., 'The economic theory of public enforcement of law' (2000) 38(1) *Journal of Economic Literature* 45.

Pratt T. C. and Cullen F. T., 'The empirical status of Gottfredson and Hirschi's General Theory of Crime: A meta-analysis' (2000) 38(3) *Criminology* 931.

Prescott A. T., Sargent J. D. and Hull J. G., 'Meta-analysis of the relationship between violent video game play and physical aggression over time' (2018) 115(40) *Proceedings of the National Academy of Sciences* 9882.

Prins S. J., 'The prevalence of mental illnesses in U.S. State prisons: A systematic review' (2014) 65(7) *Psychiatric Services* 862.

Prinz W., 'Perception and action planning' (1997) 9(2) *European Journal of Cognitive Psychology* 129.

Ptak R., Schnider A. and Fellrath J., 'The dorsal frontoparietal network: A core system for emulated action' (2017) 21(8) *Trends in Cognitive Sciences* 589.

Pylshyn Z. W. and Storm R. W., 'Tracking multiple independent targets: Evidence for a parallel tracking mechanism' (1988) 3(3) *Spatial Vision* 179.

Qian Z, Zhang D. and Wang L., 'Is aggressive trait responsible for violence? Priming effects of aggressive words and violent movies' (2013) 4(2) *Psychology* 96.

Rahwan Z., Yoeli E. and Fasolo B., 'Heterogeneity in banker culture and its influence on dishonesty' (2019) 575(7782) *Nature* 345.

Railton P., 'Moral learning: Conceptual foundations and normative relevance' (2017) 167 *Cognition* 172.

Railton P., 'The affective dog and its rational tale: Intuition and attunement' (2014) 124(4) *Ethics* 813.

Ratneshwar S., Shocker A. D. and Stewart D. W., 'Toward understanding the attraction effect: The implications of product stimulus meaningfulness and familiarity' (1987) 13(4) *Journal of Consumer Research* 520.

Ravizza M., 'Semi-compatibilism and the transfer of non-responsibility' (1994) 71(1) *Philosophical Studies* 61.

Raz J., 'Responsibility and the negligence standard' (2010) 30(1) *Oxford Journal of Legal Studies* 1.

Rennó-Costa C., Lisman J. E. and Verschure P. F. M. J., 'The mechanism of rate remapping in the dentate gyrus' (2010) 68(6) *Neuron* 1051.

Rennó-Costa C., Lisman J. E. and Verschure P. F. M. J., 'A signature of attractor dynamics in the CA3 region of the hippocampus' (2014) 10(5) *PLoS Computational Biology* e1003641.

Richeson J. A. and Shelton J. N., 'When prejudice does not pay: Effects of interracial contact on executive function' (2003) 14(3) *Psychological Science* 287.

Ridderinkhof K. R., Ullsperger M., Crone M. A. and Nieuwenhuis S., 'The role of the medial frontal cortex in cognitive control' (2004) 306(5695) *Science* 443.

Rizzolatti G., Camarda R., Fogassi L., Gentilucci M., Luppino G. and Matelli M., 'Functional organization of inferior area 6 in the macaque monkey' (1988) 71(3) *Experimental Brain Research* 491.

Rizzolatti G. and Craighero L., 'The mirror-neuron system' (2004) 27(1) *Annual Review of Neuroscience* 169.

Robinson P. H., 'Causing the conditions of one's own defense: A study in the limits of theory in criminal law doctrine' (1985) 71(1) *Virginia Law Review* 1.

Robinson P. H., 'Criminal law defences: A systematic analysis' (1982) 82(2) *Columbia Law Review* 199.

Robinson T. E. and Berridge K. C., 'The incentive sensitization theory of addiction: Some current issues' (2008) 363(1507) *Philosophical Transactions of the Royal Society: Biological Sciences* 3137.

Robinson T. E. and Berridge K. C., 'The neural basis of drug craving: An incentive-sensitization theory of addiction' (1993) 18(3) *Brain Research Reviews* 247.

Roe R. M., Busemeyer J. R. and Townsend J. T., 'Multialternative decision field theory: A dynamic connectionist model of decision making' (2001) 108(2) *Psychological Review* 370.

Rogers J., 'Dishonesty in the first LIBOR trial' (2016) 3 *Archbold Review* 7.

Rogers J., 'Necessity, private defence and the killing of Mary' (2001) (Jul) *Criminal Law Review* 515.

Rondi-Reig L., Petit G. H., Tobin C., Tonegawa S., Mariani J. and Berthoz A., 'Impaired sequential egocentric and allocentric memories in forebrain-specific-NMDA receptor knock-out mice during a new task dissociating strategies of navigation' (2006) 26(15) *Neuroscience* 4071.

Rosenbaum S. M., Billinger S. and Stieglitz N., 'Let's be honest: A review of experimental evidence of honesty and truth-telling' (2014) 45 *Journal of Economic Psychology* 181.

Ross R. R., Fabiano E. A. and Ewles C. D., 'Reasoning and rehabilitation' (1988) 32(1) *International Journal of Offender Therapy and Comparative Criminology* 29.

Rothi L. J. G., Heilman K. M. and Watson R. T., 'Pantomime comprehension and ideomotor apraxia' (1985) 48(3) *Neurosurgery & Psychiatry* 207.

Rudinow J., 'Manipulation' (1978) 88(4) *Ethics* 338.

Rushworth M. F. S., Walton M. E., Kennnerley S. W. and Bannerman D. M., 'Action sets and decisions in the medial frontal cortex' (2004) 8(9) *Trends in Cognitive Sciences* 410.

Sadurski W., 'Social justice and legal justice' (1984) 3(3) *Law and Philosophy* 329.

Sakata H., Taira M., Kusunoki M., Murata A. and Tanaka Y., 'The TINS Lecture: The parietal association cortex in depth perception and visual control of hand action' (1997) 20(8) *Trends in Neuroscience* 350.

Samejima K., Ueda Y., Doya K. and Kimura M., 'Representation of action-specific reward values in the striatum' (2005) 310(5752) *Science* 1337.

Samuels A., 'The diabetic driver' (2019) 59(4) *Medicine, Science and the Law* 282.

Sauer H., 'Education institution, automaticity and rationality in moral judgment' (2012) 15(3) *Philosophical Explorations* 255.

Sayre F. B., '*Mens rea*' (1932) 45(6) *Harvard Law Review* 974.

Scanlon T. M., 'Giving desert its due' (2013) 16(2) *Philosophical Explorations* 101.

Schacter D. L., 'Implicit memory: History and current status' (1987) 13(3) *Journal of Experimental Psychology: Learning, Memory and Cognition* 501.

Schafer J. R., 'The deterrent effect of three strikes laws' (1999) 68(1) *FBI Law Enforcement Bulletin* 6.

Schel M. A., Kühn S., Brass M., Haggard P., Ridderinkhof K. R. and Crone E. A., 'Neural correlates of intentional and stimulus-driven inhibition: a comparison' (2014) 8 *Frontiers in Human Neuroscience* 1.

Scherbaum N. and Specka M., 'Factors influencing the course of opiate addiction' (2008) 17(S1) *International Journal of Methods in Psychiatric Research* S39.

Scherberger H. and Andersen R. A., 'Target selection signals for arm reaching in the posterior parietal cortex' (2007) 27(8) *Journal of Neuroscience* 2001.

Schlam T. R., Wilson N. L., Shoda Y., Mischel W. and Ayduk O., 'Preschoolers' delay of gratification predicts their body mass 30 years later' (2013) 162(1) *Journal of Pediatrics* 90.

Schlegel A., Alexander P., Sinnott-Armstrong W., Roskies A., Tse P. U. and Wheatley T., 'Barking up the wrong free: Readiness potentials reflect processes independent of conscious will' (2013) 229(3) *Experimental Brain Research* 329.

Schmahmann J. D., 'The role of the cerebellum in cognition and emotion: Personal reflection since 1982 on the dysmetria of thought hypothesis, and its historical evolution from theory to therapy' (2010) 20(3) *Neuropsychology Review* 236.

Schmeichel B. J., 'Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control' (2007) 136(2) *Journal of Experimental Psychology* 241.

Schmeichel B. J., Vohs K. D. and Baumeister R. F., 'Intellectual performance and ego depletion: Role of the self in logical reasoning and other information processing' (2003) 85(1) *Journal of Personality and Social Psychology* 33.

Schoenbaum G., Setlow B., Saddoris M. P. and Gallagher M., 'Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala' (2003) 39(5) *Neuron* 855.

Schulz-Hardt S., Brodbeck F. C., Mojzisch A., Kerschreiter R. and Frey D., 'Group decision Making in hidden profile situations: Dissent as a facilitator for decision quality' (2006) 91(6) *Journal of Personality and Social Psychology* 1080.

Schwarz N. and Clore G. L., 'Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states' (1983) 45(3) *Journal of Personality and Social Psychology* 513.

Schweitzer M. E., Ordóñez L. D. and Douma B., 'Goal setting as a motivator of unethical behavior' (2004) 47(3) *Academy of Management Journal* 422.

Scott D., 'The effect of video games on feelings of aggression' (1995) 129(2) *Journal of Psychology* 121.

Seeyave D. M., Coleman S., Appugliese D., Corwyn R. F., Bradley R. H., Davidson N. S., Kaciroti N. and Lumeng J. C., 'Ability to delay gratification at age 4 years and risk of overweight at age 11 years' (2009) 163(4) *Archives of Pediatrics and Adolescent Medicine* 303.

Segal S. J., 'The priming of association test responses: Generalizing the phenomenon' (1967) 6(2) *Journal of Verbal Learning and Verbal Behavior* 216.

Segal S. J. and Cofer C. N., 'The effect of recency and recall on word association' (1960) 15 *American Psychologist* 451.

Seidler R. D., Noll D. C. and Thiers G., 'Feedforward and feedback processes in motor control' (2004) 22(4) *NeuroImage* 1775.

Selfe D. W., 'Rape: *Mens rea* and reasonable belief' (2013) 214 *Criminal Lawyer* 3.

Sell A., Tooby J. and Cosmides L., 'Formidability and the logic of human anger' (2009) 106(35) *Proceedings of the National Academy of Sciences* 15073.

Seneviratne M., 'Carry on Caldwell' (2003) 12(1) *Nottingham Law Journal* 36.

Serences J. T. and Yantis S., 'Selective visual attention and perceptual coherence' (2006) 10(1) *Trends in Cognitive Sciences* 38.

Shah J. Y., Friedman R. and Kruglanski A. W., 'Forgetting all else: On the antecedents and consequences of goal shielding' (2002) 83(6) *Journal of Personality and Social Psychology* 1261.

Shariff A. F. and Norenzayan A., 'God is watching you: Priming god concepts increases prosocial behavior in an anonymous economic game' (2007) 18(9) *Psychological Science* 803.

Sharot T., Felming S. M., Yu X., Koster R. and Dolan R. J., 'Is choice-induced preference change long lasting?' (2012) 23(10) *Psychological Science* 1123.

Shavitt S., 'The role of attitude objects in attitude functions' (1990) 26(2) *Journal of Experimental Social Psychology* 124.

Sheeran P., Webb T. L. and Gollwitzer P. M., 'The interplay between goal intentions and implementation intentions' (2005) 31(1) *Personality and Social Psychology Bulletin* 87.

Shergill S., Samson G., Bays P. M., Frith C. D. and Wolpert D. M., 'Evidence for sensory prediction deficits in schizophrenia' (2005) 162(12) *American Journal of Psychiatry* 2384.

Sherman S. J., Mackie D. M. and Driscoll D. M., 'Priming and the differential use of dimensions in evaluation' (1990) 16(3) *Personality and Social Psychology Bulletin* 405.

Sherry J. L., 'The effects of violent video games on aggression: A meta-analysis' (2006) 27(3) *Human Communications Research* 409.

Shibasaki H., 'Cortical activities associated with voluntary movements and involuntary movements' (2012) 123(2) *Clinical Neurophysiology* 229.

Shibasaki H. and Hallet M., 'What is the bereitschaftspotential?' (2006) 117(11) *Clinical Neurophysiology* 2341.

Shih M., Ambady N., Richeson J. A., Fujita K. and Gray H. M., 'Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation' (2002) 83(3) *Journal of Personality and Social Psychology* 638.

Shmuelof L. and Zohary E., 'Dissociation between ventral and dorsal fMRI activation during object and action recognition' (2005) 47(3) *Neuron* 457.

Shoda Y., Mischel W. and Peake P. K., 'Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions' (1990) 26(6) *Developmental Psychology* 978.

Simester A. P., 'Mistakes in defence' (1992) 12(2) *Oxford Journal of Legal Studies* 295.

Simester A. P., 'Moral certainty and the boundaries of intention' (1996) 16(3) *Oxford Journal of Legal Studies* 445.

Simonson I., 'Choice based on reasons: The case of attraction and compromise effects' (1989) 16(2) *Journal of Consumer Research* 158.

Simonson I. and Tversky A., 'Choice in context: Tradeoff contrast and extremeness aversion' (1992) 29(3) *Journal of Marketing Research* 281.

Singer R. G., 'The resurgence of *mens rea*: I – Provocation, emotional disturbance, and the model penal code' (1986) 27(2) *Boston College Law Review* 243.

Singer R. G., 'The resurgence of *mens rea*: III – The rise and fall of strict criminal liability' (1989) 30(2) *Boston College Law Review* 337.

Singleton N., Farrell N. and Meltzer H., 'Substance misuse among prisoners in England and Wales' (2003) 15(1-2) *International Review of Psychiatry* 150.

Sirigu A., Daprati E., Ciancia S., Giraux P., Nighoghossian N., Posada A. and Haggard P., 'Altered awareness of voluntary action after damage to the parietal cortex' (2003) 7(1) *Nature Neuroscience* 80.

Sirigu A., Daprati E., Pradat-Diehl P., Franck N. and Jeannerod M., 'Perception of self-generated movement following left parietal lesion' (1999) 122(10) *Brain* 1867.

Sjöberg L., 'Choice frequency and similarity' (1977) 18(1) *Scandinavian Journal of Psychology* 103.

Slater J., 'Making sense of self-defence' (1996) 5(2) *Nottingham Law Journal* 140.

Smetana J. G., 'Social-cognitive development: Domain distinctions and coordinations' (1983) 3(2) *Developmental Review* 131.

Smith A. T. H., 'Judicial law making in the criminal law' (1984) 100(1) *Law Quarterly Review* 46.

Smith E. R. and Branscombe N. R., 'Procedurally mediated social inference: The case of category accessibility effects' (1987) 23(5) *Journal of Experimental Social Psychology* 361.

Smith J. C., 'Case commentary: *R v Caldwell*' (1981) *Criminal Law Review* 392.

Smith J. C., 'Case commentary: *R v Woollin*' (1998) (Dec) *Criminal Law Review* 890.

Smith J. C., 'The use of force in public or private defence and Article 2' (2002) (Dec) *Criminal Law Review* 958.

Smith P., Gendreau P. and Swatz K., 'Validating the principles of effective intervention: A systematic review of the contributions of meta-analysis in the field of corrections' (2009) 4(2) *Victims and Offenders* 148.

Solomon R. L. and Corbit J. D., 'An opponent-process theory of motivation' (1974) 81(2) *Psychological Review* 119.

Somerville L. H., Hare T. and Casey B. J., 'Frontostriatal maturation predicts cognitive control failure to appetitive cues in adolescents' (2011) 23(9) *Journal of Cognitive Neuroscience* 2123.

Soon C. S., Brass M., Heinze H.-J. and Haynes J.-D., 'Unconscious determinants of free decisions in the human brain' (2008) 11(5) *Nature Neuroscience* 543.

Soon C. S., He A. H., Bode S. and Haynes J.-D., 'Predicting free choices for abstract intentions' (2013) 110(15) *Proceedings of the National Academy of Sciences* 6217.

Speight A., '"Listening to reason": The role of persuasion in Aristotle's account of praise, blame, and the voluntary' (2005) 38(3) *Philosophy & Rhetoric* 213.

Sperber D., 'An evolutionary perspective on testimony and argumentation' (2001) 29(1/2) *Philosophical Topics* 401.

Sperber D., Clément F., Heintz C., Mascaro O., Mercier H., Origgi G. and Wilson D., 'Epistemic vigilance' (2010) 25(4) *Mind & Language* 359.

Srull T. K. and Wyer R. S., 'The role of category accessibility in the interpretation of information about persons: Some determinants and implications' (1979) 37(10) *Journal of Personality and Social Psychology* 1660.

Stannard J. E., 'From Andrews to Seymour and back again' (1996) 47(1) *Northern Ireland Legal Quarterly* 1.

Stannard J. E., 'Murder and the ruthless risk-taker' (2008) 8(2) *Oxford University Commonwealth Law Journal* 137.

Stasson M. F., Kameda T., Parks C. D., Zimmerman S. K. and Davis J. H., 'Effects of assigned group consensus requirement on group problem solving and group members' learning' (1991) 54(1) *Social Psychology Quarterly* 25.

Steel A., 'The meanings of dishonesty' (2009) 38(2) *Common Law World Review* 103.

Stetson C., Cui X., Montague P. R. and Eagleman D. M., 'Motor-sensory recalibration leads to an illusory reversal of action and sensation' (2006) 51(5) *Neuron* 651.

Stewart B. D. and Payne B. K., 'Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control' (2008) 34(10) *Personality and Social Psychology Bulletin* 1332.

Stewart B. M., Baugh L. A., Gallivan J. P. and Flanagan J. R., 'Simultaneous encoding of the direction and orientation of potential targets during reach planning: Evidence of multiple competing reach plans' (2013) 110(4) *Journal of Neurophysiology* 807.

Stewart B. M., Gallivan J. P., Baugh L. A. and Flanagan J. R., 'Motor, not visual, encoding of potential reach targets' (2014) 24(19) *Current Biology* R953.

Stewart H., 'Legality and morality in H. L. A. Hart's theory of criminal law' (1999) 52(1) *SMU Law Review* 201.

Stolzenberg D. S. and Numan M., 'Hypothalamic interaction with the mesolimbic DA system in the control of the maternal and sexual behaviors in rats' (2011) 35(3) *Neuroscience & Biobehavioral Reviews* 826.

Storms L. H., 'Apparent backward association: A situational effect' (1958) 55(4) *Experimental Psychology* 390.

Strack F. and Deutsch R., 'Reflective and impulsive determinants of social behavior' (2004) 8(3) *Personality and Social Psychology Review* 220.

Strandberg T., Olson J. A., Hall L., Woods A. and Johansson P., 'Depolarizing American voters: Democrats and Republicans are equally susceptible to false attitude feedback' (2020) 15(2) *PLoS ONE* e0226799.

Strandberg T., Sivén D., Hall L., Johansson P. and Pärnamets P., 'False beliefs and confabulation can lead to lasting changes in political attitudes' (2018) 147(9) *Journal of Experimental Psychology* 1382.

Strang H., Sherman L. W., Maro-Wilson E., Woods D. and Ariel B., 'Restorative justice conferencing (RJC) using face-to-face meetings of offenders and victims: Effects on offender recidivism and victim satisfaction. A systematic review' (2013) 12(1) *Campbell Systematic Reviews* 1.

Street M. D., Douglas S. C., Geiger S. W. and Martinko M. J., 'The impact of cognitive expenditure on the ethical decision-making process: The cognitive elaboration model' (2001) 86(2) *Organizational Behavior and Human Decision Processes* 256.

Stucke T. S. and Baumeister R. F., 'Ego depletion and aggressive behavior: Is the inhibition of aggression a limited resource?' (2006) 36(1) *European Journal of Social Psychology* 1.

Sugrue L. P., Corrado G. S. and Newsome W. T., 'Matching behavior and the representation of value in the parietal cortex' (2004) 304(5678) *Science* 1782**.**

Sullivan G. R. and Simester A. P., 'Judging dishonesty' (2020) 136(Oct) *Law Quarterly Review* 523.

Summers J. S., '*Post hoc ergo propter hoc*: some benefits of rationalization' (2017) 20(1) *Philosophical Explorations* 21.

Sutcliffe J. G. and de Lecea L., 'The hypocretins: Setting the arousal threshold' (2002) 3(5) *Nature Reviews Neuroscience* 339.

Synofzik M., Vosgerau G. and Voss M., 'The experience of agency: An interplay between prediction and postdiction' (2013) 4(127) *Frontiers in Psychology* 1.

Taggart C. P., 'Retributivism, agency, and the voluntary act requirement' (2016) 36(3) *Pace Law Review* 645.

Taggart C. P., 'Retributivism, ultimate responsibility, and agent causalism' (2019) 54(3) *Tulsa Law Review* 441.

Takashima S., Cravo A. M., Sameshima K. and Ramos R. T., 'The effect of conscious intention to act on the bereitschaftspotential' (2018) 236(9) *Experimental Brain Research* 2287.

Takashima S., Najman F. A. and Ramos R. T., 'Disruption of volitional control in obsessive-compulsive disorder: Evidence from the bereitschaftspotential' (2019) 290 *Psychiatry Research: Neuroimaging* 30.

Takashima S., Ogawa C. Y., Najman F. A. and Ramos R. T., 'The volition, the mode of movement selection and the readiness potential' (2020) 238(10) *Experimental Brain Research* 2113.

Talbot J. and Riley C., 'No one knows: Offenders with learning difficulties and learning disabilities' (2007) 35(3) *British Journal of Learning Disabilities* 154.

Tangney J. P., Baumeister R. F. and Boone A. L., 'High self-control predicts good adjustment, less pathology, better grades, and interpersonal success' (2004) 72(2) *Journal of Personality* 271.

Tanné-Gariépy J., Rouiller E. M. and Boussaoud D., 'Parietal inputs to dorsal versus ventral premotor areas in the macaque monkey: evidence for largely segregated visuomotor pathways' (2002) 145(1) *Experimental Brain Research* 91.

Tanner R. J., Ferraro R., Chartrand T. L., Bettman J. F. and van Baaren R., 'Of chameleons and consumption: The impact of mimicry on choice and preferences' (2008) 34(6) *Journal of Consumer Research* 754.

Taylor J. L. and McCloskey D. I., 'Triggering of preprogrammed movements as reactions to masked stimuli' (1990) 63(3) *Journal of Neurophysiology* 439.

Teixeira S., Machado S., Velasques B., Sanfim A., Minc D., Peressutti C., Bittencourt J., Budde H., Cagy M., Anghinah R., Basile L. F., Piedade R., Ribeiro P., Diniz C., Cartier C., Gongora M., Silva F., Manaia F. and Silva J. G., 'Integrative parietal cortex processes: Neurological and psychiatric aspects' (2014) 338(1-2) *Journal of the Neurological Sciences* 12.

Temkin J. and Ashworth A., 'The Sexual Offences Act 2003: (1) Rape, sexual assaults and the problems of consent' (2004) (May) *Criminal Law Review* 328.

Thaut M. H., 'Neural basis of rhythmic timing networks in the human brain' (2003) 999(1) *Annals of the New York Academy of Sciences* 364.

Thomas M., '"Reasonable cause to suspect": In the absence of knowledge and actual suspicion' (2018) 82(6) *Journal of Criminal Law* 423.

Thomas M. and Pegg S., 'Clarifying the applicable test for dishonesty and modifying *stare decisis*, but otherwise a missed opportunity' (2020) 84(4) *Journal of Criminal Law* 385.

Thompson V. A., Evans J. St. B. T. and Handley S. J., 'Persuading and dissuading by conditional argument' (2005) 53(2) *Journal of Memory and Language* 238.

Thornton M. T., 'Making sense of *Majewski*' (1980-81) 23(4) *Criminal Law Quarterly* 464.

Tolman E. C., 'Cognitive maps in rats and men' (1948) 55(4) *Psychological Review* 189.

Tomasello M., Carpenter M., Call J., Behne T. and Moll H., 'Understanding and sharing intentions: The origins of cultural cognition' (2005) 28(5) *Behavioral and Brain Sciences* 675.

Tong L. S. J. and Farrington D. P., 'How effective is the "Reasoning and Rehabilitation" programme in reducing reoffending? A meta-analysis of evaluations in four countries' (2006) 12(1) *Psychology, Crime and Law* 3.

Tormala Z. L. and Petty R. E., 'What doesn't kill me makes me stronger: The effects of resisting persuasion on attitude certainty' (2002) 83(6) *Journal of Personality and Social Psychology* 1298.

Tremblay L. and Schultz W., 'Relative reward preference in primate orbitofrontal cortex' (1999) 398(6729) *Nature* 704.

Trouche E., Johansson P., Hall L. and Mercier H., 'The selective laziness of reasoning' (2015) 40(8) *Cognitive Science* 2122.

Tsujimoto S., Genovesio A. and Wise S. P., 'Transient neuronal correlations underlying goal selection and maintenance in prefrontal cortex' (2008) 18(12) *Cerebral Cortex* 2748.

Tulving E., 'Episodic memory: From mind to brain' (2002) 53(1) *Annual Review of Psychology* 1.

Tulving E. and Pearlstone Z., 'Availability versus accessibility of information in memory for words' (1966) 5(4) *Journal of Verbal Learning and Verbal Behavior* 381.

Tulving E. and Schacter D. L., 'Priming and human memory systems' (1990) 247(4940) *Science* 301.

Tulving E., Schacter D. L. and Stark H. A., 'Priming effects in word-fragment completion are independent of recognition memory' (1982) 8(4) *Journal of Experimental Psychology: Learning, Memory and Cognition* 336.

Tunik E., Frey S. H. and Grafton S. T., 'Virtual lesions of the anterior intraparietal area disrupt goal-dependent on-line adjustments of grasp' (2005) 8(4) *Nature Neuroscience* 505.

Tunik E., Rice N. J., Hamilton A. F. and Grafton S. T., 'Beyond grasping: representation of action in human anterior intraparietal sulcus' (2007) 36(Supp. 2) *NeuroImage* T77.

Turiel E., Hildebrandt C., Wainryb C. and Saltzstein H. D., 'Judging social issues: Difficulties, inconsistencies, and consistencies' (1991) 56(2) *Monographs of the Society for Research in Child Development* 1.

Turin L., 'A spectroscopic mechanism for primary olfactory reception' (1996) 21(6) *Chemical Senses* 773.

Turner J. W. C., '*Mens rea* and motorists' (1933) 5 *Cambridge Law Journal* 61.

Turner M. and Coltheart M., 'Confabulation and delusion: a common monitoring framework' (2010) 15(1) *Cognitive Neuropsychology* 346.

Tusche A., Bode S. and Haynes J.-D., 'Neural responses to unattended products predict later consumer choices' (2010) 30(23) *The Journal of Neuroscience* 8024.

Tversky A., 'Elimination by aspects: A theory of choice' (1972) 79(4) *Psychological Review* 281.

Tversky A. and Kahneman D., 'Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment' (1983) 90(4) *Psychological Review* 293.

Tversky A. and Simonson I., 'Context-dependent preferences' (1993) 39(10) *Management Science* 1179.

Uhlmann L., Pazen M., van Kemenade B. M., Kircher T. and Straube B., 'Neural correlates of self-other distinction in patients with schizophrenia spectrum disorders: The role of agency and hand identity' (2021) 47(5) *Schizophrenia Bulletin* 1399.

Vaziri A. and Plenio M. B., 'Quantum coherence in ion channels: resonances, transport and verification' (2010) 12(8) *New Journal of Physics* 085001.

Vazsonyi A. T., Mikuška J. and Kelley E. L., 'It's time: A meta-analysis on the self-control – deviance link' (2017) 48 *Journal of Criminal Justice* 48.

Verbruggen F. and Logan G. D., 'Automaticity of cognitive control: Goal priming in response-inhibition paradigms' (2009) 35(5) *Journal of Experimental Psychology* 1381.

Vernon-Feagans L., Willoughby M., Garrett-Peters P. and Family Life Project Key Investigators, 'Predictors of behavioral regulation in kindergarten: Household chaos, parenting, and early executive functions' (2016) 52(3) *Developmental Psychology* 430.

Verschure P. F. M. J., Pennartz C. M. A. and Pezzulo G., 'The why, what, where, when and how of goal-directed choice: neuronal and computational principles' (2014) 369(1655) *Philosophical Transactions of the Royal Society: Biological Sciences* 20130483.

Vingerhoets G., 'Contribution of the posterior parietal cortex in reaching, grasping, and using objects and tools' (2014) 5 *Frontiers in Psychology* 151.

Virgo G., 'Laundering conspiracy' (206) 65(3) *Cambridge Law Journal* 482.

Virgo G., 'Cheating and dishonesty' (2018) 77(1) *Cambridge Law Journal* 18.

Vohs K. D., Baumeister R. F. and Ciarocco N., 'Self-regulation and self-presentation: Regulatory resource depletion impairs impression management and effortful self-presentation depletes regulatory resources' (2005) 88(4) *Journal of Personality and Social Psychology* 632.

Vohs K. D. and Heatherton T. F., 'Self-regulatory failure: A resource-depletion approach' (2000) 11(3) *Psychological Science* 249.

Vohs K. D. and Schooler J. W., 'The value of believing in free will: encouraging a belief in determinism increases cheating' (2008) 19(1) *Psychological Science* 49.

Volkow N. D., Fowler J. S., Wang G.-J., Baler R. and Telang F., 'Imaging dopamine's role in drug abuse and addiction' (2009) 56(1) *Neuropharmacology* 3.

Volkow N. D., Fowler J. S. and Wang G.-J., 'Role of dopamine in drug reinforcement and addiction in humans: Results from imaging studies' (2002) 13(5) *Behavioural Pharmacology* 355.

Vorberg D., Matler U., Heinecke A., Schmidt T. and Schwarzbach J., 'Different time courses for visual perception and action priming' (2003) 100(10) *Proceedings of the National Academy of Sciences* 6275.

Voss M., Moore J., Hauser M., Gallinat J., Heinz A. and Haggard P., 'Altered awareness of action in schizophrenia: A specific deficit in predicting action consequences' (2010) 133(10) *Brain* 3104.

Vranken M., 'Duty to rescue in civil law and common law: Les extremes se touchent?' (1998) 47(4) *International and Comparative Law Quarterly* 934.

Wager T. D., Sylvester C.-Y. C., Lacey S. C., Nee D. E., Franklin M. and Jonides J., 'Common and unique components of response inhibition revealed by fMRI' (2005) 27(2) *NeuroImage* 323.

Wake N., 'Political rhetoric or principled reform of loss of control? Anglo-Australian perspectives on the exclusionary conduct model' (2013) 77(6) *Journal of Criminal Law* 512.

Wall S., 'Enforcing morality' (2013) 7(3) *Criminal Law and Philosophy* 455.

Wallis J. D., Anderson K. C. and Miller E. K., 'Single neurons in prefrontal cortex encode abstract rules' (2001) 411(6840) *Nature* 953.

Walsh E. and Haggard P., 'Action, prediction, and temporal awareness' (2013) 142(2) *Acta Psychologica* 220.

Walsh E., Kühn S., Brass M., Wenke D. and Haggard P., 'EEG activations during intentional inhibition of voluntary action: An electrophysiological correlate of self-control?' (2010) 48(2) *Neuropsychologia* 619.

Walters M. A., Paterson J. L., McDonnell L. and Brown R., 'Group identity, empathy and shared suffering: Understanding the "community" impacts of anti-LGBT and Islamophobic hate crimes' (2019) 26(2) *International Review of Victimology* 143.

Walton M. E., Devlin J. T. and Rushworth M. F. S., 'Interactions between decision making and performance monitoring within prefrontal cortex' (2004) 7(11) *Nature Neuroscience* 1259.

Wang X.-J., 'Probabilistic decision making by slow reverberation in cortical circuits' (2002) 36(5) *Neuron* 955.

Wang Y., Wang G., Chen Q. and Li L., 'Depletion, moral identity, and unethical behavior: Why people behave unethically after self-control exertion' (2017) 56 *Consciousness and Cognition* 188.

Ward A. R., 'Making some sense of self-induced intoxication' (1986) 45(2) *Cambridge Law Journal* 247.

Ware A., 'The conception of manipulation: Its relation to democracy and power' (1981) 11(2) *British Journal of Political Science* 163.

Wasik M. and Thompson M., '"Turning a blind eye" as constituting *mens rea*' (1981) 32 *Northern Ireland Law Quarterly* 328.

Wasserman D. and Johnston J., 'Seeing responsibility: Can neuroimaging teach us anything about moral and legal responsibility' (2014) 44(s2) *Hastings Center Report* S37.

Wasserstrom R. A., 'H. L. A. Hart and the doctrines of *mens rea* and criminal responsibility' (1967) 35(1) *University of Chicago Law Review* 92.

Watts T. W., Duncan G. J. and Quan H., 'Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes' (2018) 29(7) *Psychological Science* 1159.

Wegner D. M., 'The mind's best trick: how we experience conscious will' (2003) 7(2) *Trends in Cognitive Sciences* 65.

Wegner D. M. and Wheatley T., 'Apparent mental causation: Sources of the experience of will' (1999) 54(7) *American Psychologist* 480.

Wegner D. M., Schneider D. J., Carter S. R. and White T. L., 'Paradoxical effects of thought suppression' (1987) 53(1) *Journal of Personality and Social Psychology* 5.

Weiskrantz L., 'Blindsight revisited' (1996) 6(2) *Current Opinion in Neurobiology* 215.

Weiskrantz L., Elliott J. and Darlington C. D., 'Preliminary observations on tickling oneself' (1971) 230(5296) *Nature* 598.

Welsh D. T. and Ordóñez L. D., 'The dark side of consecutive high performance goals: Linking goal setting, depletion, and unethical behavior' (2014) 123(2) *Organizational Behavior and Human Decision Processes* 79.

Wentura D. and Rothermund K., 'Priming is not priming is not priming' (2014) 32 (Supp) *Social Cognition* 47.

Wheatley T. and Haidt J., 'Hypnotic disgust makes moral judgments more severe' (2005) 16(10) *Psychological Science* 780.

Wheeler S. C., DeMarree K. G. and Petty R. E, 'Understanding prime-to-behavior effects: Insights from the active-self account' (2014) 32(Supp) *Social Cognition* 109.

Wheeler S. C., DeMarree K. G. and Petty R. E., 'Understanding the role of the self in prime-to-behavior effects: The active-self account' (2007) 11(3) *Personality and Social Psychology Review* 234.

Wheeler S. C., Jarvis W. B. G. and Petty R. E., 'Think unto others: The self-destructive impact of negative racial stereotypes' (2001) 37(2) *Journal of Experimental Social Psychology* 173.

Widerker D., 'Frankfurt on "ought implies can" and alternative possibilities' (1991) 51(4) *Analysis* 222.

Williams G., 'Necessity: Duress of circumstances of moral involuntariness?' (2014) 43(1) *Common Law World Review* 1.

Williams G., 'Oblique intention' (1987) 46(3) *Cambridge Law Journal* 417.

Williams G., 'Recklessness redefined' (1981) 40(2) *Cambridge Law Journal* 252.

Williams G., 'The *mens rea* for murder: Leave it alone' (1989) 105(Jul) *Law Quarterly Review* 387.

Williams G., 'The standard of honesty' (1983) 133 *New Law Journal* 636.

Williams G., 'The unresolved problem of recklessness' (1988) 8(1) *Legal Studies* 74.

Wilson D. B., Bouffard L. A. and Mackenzie D. L., 'A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders' (2005) 32(2) *Criminal Justice and Behavior* 172.

Wilson M. A. and McNaughton B. L., 'Reactivation of hippocampal ensemble memories during sleep' (1994) 265(5172) *Science* 676.

Wilson M. L., Britton N. F. and Franks N. R., 'Chimpanzees and the mathematics of battle' (2002) 269(1496) *Proceedings of the Royal Society: Biological Sciences* 1107.

Wilson T. D. and Brekke N., 'Mental contamination and mental correction: Unwanted influences on judgments and evaluations' (1994) 116(1) *Psychological Bulletin* 117.

Wilson T. D., Lindsey S. and Schooler T. Y., 'A modal of dual attitudes' (2000) 107(1) *Psychological Review* 101.

Wilson W., Ebrahim I., Fenwick P. and Marks R., 'Violence, sleepwalking and the criminal law: Part 2: The legal aspects' (2005) (Aug) *Criminal Law Review* 614.

Windzio M., 'Is there a deterrent effect of pains of imprisonment? The impact of "social costs" of first incarceration on the hazard rate of recidivism' (2006) 8(3) *Punishment and Society* 341.

Wisniewski D., Goschke T. and Haynes J.-D., 'Similar coding of freely chosen and externally cued intentions in a fronto-parietal network' (2016) 134 *NeuroImage* 450.

Withey C., 'Loss of control, loss of opportunity?' (2011) 4 *Criminal Law Review* 263.

Woike B., Lavezzary E. and Barsky J., 'The influence of implicit motives on memory processes' (2001) 81(5) *Journal of Personality and Social Psychology* 935.

Wolpert D. M., 'Computational approaches to motor control' (1997) 1(6) *Trends in Cognitive Sciences* 209.

Wolpert D. M. and Ghahramani Z., 'Computational principles of movement neuroscience' (2000) 3(supp) *Nature Neuroscience* 1212.

Wolpert D. M., Miall R. C. and Kawato M., 'Internal models in the cerebellum' (1998) 2(9) *Trends in Cognitive Sciences* 338.

Wong A. L., Haith A. M. and Krakauer J. W., 'Motor planning' (2015) 21(4) *Neuroscientist* 385.

Wood E. R., Dudchenko P. A. and Eichenbaum H., 'The global record of memory in hippocampal neuronal activity' (1999) 397(6720) *Nature* 613.

Works E., 'The prejudice-interaction hypothesis from the point of view of the negro minority group' (1961) 67(1) *American Journal of Sociology* 47.

Wortley N., 'Reasonable belief in consent under the Sexual Offences Act 2003' (2013) 77(3) *Journal of Criminal Law* 184.

Wu A., 'Going full circle: Gender and the "loss of control" defence under the Coroners and Justice Act 2009' (2019) 1(1) *Rule of Law Journal* 46.

Wu Y.-W., Zhong L.-L., Ruan Q.-N., Liang J. and Yan W.-J., 'Can priming legal consequences and the concept of honesty decrease cheating during examinations?' (2020) 10 *Frontiers in Psychology* 2887.

Yang G. S., Huesmann L. R. and Bushman B. J., 'Effects of playing a violent video game as male versus female avatar on subsequent aggression in male and female players' (2014) 40(6) *Aggressive Behavior* 537.

Yang T. and Shadlen M.N., 'Probabilistic reasoning by neurons' (2007) 447(7148) *Nature* 1075.

Yao M. Z., Mahood C. and Linz D., 'Sexual priming, gender stereotyping, and likelihood to sexually harass: Examining the cognitive effects of playing a sexually-explicit video game' (2010) 62(1/2) *Sex Roles* 77.

Young D., 'Excuses and intelligibility in criminal law' (2004) 79(1) *University of New Brunswick Law Journal* 79.

Yu H., McCuller L., Tse M., Kijbunchoo N., Barsotti L., Mavalvala N. and members of the LIGO Scientific Collaboration, 'Quantum correlations between lights and the kilogram-mass mirrors of LIGO' (2020) 583(7814) *Nature* 43.

Zapparoli L., Porta M. and Paulesu E., 'The anarchic brain in action: The contribution of task-based fMRI studies to the understanding of Gilles de la Tourette syndrome' (2015) 28(6) *Current Opinion in Neurology* 604.

Zapparoli L., Seghezzi S. and Paulesu E., 'The what, the when, and the whether of intentional action in the brain: A meta-analytical review' (2017) 11 *Frontier in Human Neuroscience* 1.

Zedner L., 'Preventative justice or pre-punishment? The case of control orders' (2007) 60(1) *Current Legal Problems* 174.

Zendle D., Cairns P. and Kudenko D., 'No priming in video games' (2017) 78 *Computers in Human Behavior* 113.

Zhang J., Hughes L. E. and Rowe J. B., 'Selection and inhibition mechanisms for human voluntary action decisions' (2012) 63(1) *NeuroImage* 392.

Zhang J., Kriegeskorte N., Carlin J. D. and Rowe J. B., 'Choosing the rules: Distinct and overlapping frontoparietal representations of task rules for perceptual decisions' (2013) 33(29) *Journal of Neuroscience* 11852.

Zhang Q., Tian J.J., Cao J., Zhang D.-J. and Rodkin P., 'Exposure to weapon pictures and subsequent aggression during adolescence' (2016) 90 *Personality and Individual Differences* 113.

## ***Reports, White Papers, Websites etc.***

Alcohol Health Alliance UK, 'Measuring up: The state of the nation' (Alcohol Health Alliance UK 2017) < https://ahauk.org/wp-content/uploads/2017/12/7119-AHA-10-year-anniversary-report.pdf> accessed 5 October 2022.

Beard J., Sturge G., Lalic M. and Holland S., 'General debate on the cost and effectiveness of sentences under 12 months and consequences for the prison population' (House of Commons Library, Debate pack CDP-2019-0063, March 2019).

Butler Committee, *Report of the Committee on Mentally Abnormal Offenders* (Cmnd 6244, 1975).

California Legislative Analysts' Office, 'A primer: Three strikes – The impact after more than a decade' (*Legislative Analyst's Office,* October 2005) <https://lao.ca.gov/2005/3_strikes/3_strikes_102005.htm> accessed 19 October 2022.

Cohn A. and Maréchal M. A., 'Priming in economics' (University of Zurich Department of Economics, Working paper series no. 226, 2016).

Deckers T., Falk A., Kosse F., Pinger P. and Schildberg-Hörisch H., 'Socio-economic status and inequalities in children's IQ and economic preferences' (IZA Institute of Labor Economics, Discussion paper no. 11158, November 2017).

Eagleman, 'The brain on trial' (*The Atlantic*, Jul/Aug 2011) <https://www.theatlantic.com/magazine/archive/2011/07/the-brain-on-trial/308520/> accessed 15 January 2022.

Edis J., '*R v Long, Bowers, Cole and King* – Sentencing remarks' (31st July 2020).

Engel C., 'Low self-control as a source of crime: A meta-study' (Reprints of the Max Planck Institute for Research on Collective Goods, Bonn 2012/4).

Falk A. and Kosse F., 'Early childhood environment, breastfeeding and the formation of preferences' (SOEP Papers on multidisciplinary panel data research 882-2016).

Felson M. and Clarke R. V., 'Opportunity makes the thief: Practical theory for crime prevention' (Home Office Policing and Reducing Crime Unit, Police research series paper 98, 1998).

Friehe T. and Schildberg-Hörisch H., 'Self-control and crime revisited: Disentangling the effect of self-control on risk taking and antisocial behavior' (DICE Discussion paper no. 264, 2017).

Gendreua P., Goggin C. and Cullen F. T., 'The effects of prison sentences of recidivism' (Public Works and Government Services Canada Report, 1999-3).

Gormley J., Hamilton M. and Belton I., 'The effectiveness of sentencing options on reoffending' (UK Sentencing Council 2022).

Gowers E., *Report of the Royal Commission on Capital Punishment* (Cmd 8932, 1953).

Home Office, 'FIRE0402: Fatalities and non-fatal casualties in deliberate fires by fire and rescue authority' (Home Office Fire Statistics Data Tables, November 2020) <https://www.gov.uk/government/statistical-data-sets/fire-statistics-data-tables#deliberate-fires-attended> accessed 12/01/2021.

Home Office, *Protecting the Public: Strengthening Protection Against Sex Offenders and Reforming the Law on Sexual Offences* (Cmnd 5668, 2002).

Home Office, *Setting the Boundaries: Reforming the law on sex offences* (Home Office 2000).

House of Commons Committee of Public Accounts, *Mental Health in Prisons* (HC 400, Eighth Report of Session 2017-19).

House of Commons Home Affairs Committee, *Sexual Offences Bill* (HC 639, Fifth Report of Session 2002-03).

Institute of Alcohol Studies, 'Crime and social impacts' (IAS 2019) <http://www.ias.org.uk/Alcohol-knowledge-centre/Crime-and-social-impacts.aspx#_ftn2> accessed 10 November 2020.

Joint Parliamentary Committee on Human Rights, *Legislative Scrutiny: Sixth Progress Report* (HL 134, HC 955, Fourteenth Report of Session 2005-06).

Law Commission, *Conspiracy and Attempts: A Consultation Paper* (Law Com No 183, 2007).

Law Commission, *Criminal Liability: Insanity and Automatism – A Discussion Paper* (Law Commission 2013).

Law Commission, *Fraud: Report on a Reference under Section 3(1)(e) of the Law Commissions Act 1965* (Law Com No. 276, 2002).

Law Commission, *Legislating the Criminal Code: Offences Against the Person and General Principles* (Law Com No 218, 1989).

Law Commission, *Partial Defences to Murder* (Law Com No 290, 2004).

Light M., Grant E. and Hopkins K., *Gender Differences in Substance Misuse and Mental Health Amongst Prisoners: Results from the Surveying Prisoner Crime Reduction (SPCR) Longitudinal Cohort Study of Prisoners* (Ministry of Justice 2013).

Mikhail J. M., 'Aspects of the theory of moral cognition: Investigating intuitive knowledge of the prohibition of intentional battery and the principle of double effect' (Georgetown University Law Center, Working paper no. 762385, 2002).

Mikhail J. M., 'Rawls' linguistic analogy: A study of the "generative grammar" model of moral theory described by John Rawls in *A Theory of Justice*' (DPhil thesis, Cornell University 2000).

Office for National Statistics, 'The nature of violent crime in England and Wales: year ending March 2018' (ONS 2018) <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/thenatureofviolentcrimeinenglandandwales/yearendingmarch2018> accessed 10 November 2020.

Populus, *Dignity in Dying Poll* (2015) [Online] <http://www.populus.co.uk/wp-content/uploads/2015/12/DIGNITY-IN-DYING-Populus-poll-March-2015-data-tables-with-full-party-crossbreaks.compressed.pdf> accessed 20th October 2020.

Public Health England, *Psychosis Data Report: Describing variation in numbers of people with psychosis and their access to care in England* (Crown Copyright 2016).

Robinson D., 'The impact of cognitive skills training on post-release recidivism among Canadian federal offenders' (Correctional Service Canada, Research report no. R-41, 1995).

Smits J. M., 'The good Samaritan in European private law: On the perils of principles without a programme and a programme for the future' (Inaugural lecture, Maastricht University, 19 May 2000) <https://core.ac.uk/download/pdf/231273801.pdf> accessed 16 October 2022.

Vohs K. D., Baumeister R. F., Twenge J. M., Schmeichel B. J., Tice D. M. and Crocker J., 'Decision fatigue exhausts self-regulatory resources – But so does accommodating to unchosen alternative' (2005) (unpublished) <https://www.researchgate.net/publication/237738528_Decision_Fatigue_Exhausts_Self-Regulatory_Resources_-_But_So_Does_Accommodating_to_Unchosen_Alternatives> accessed 29 January 2021.

Williams K., Poyser J. and Hopkins K., *Accommodation, Homelessness and Reoffending of Prisoners: Results from the Surveying Prisoner Crime Reduction (SPCR) Survey* (Ministry of Justice 2012).

## ***Jurisprudence***

*Airedale NHS Trust v Bland* [1993] AC 789.

*Albert v Lavin* [1982] AC 546.

*Alcock v Chief Constable of South Yorkshire Police* [1992] 1 AC 310.

*Andrews v Director of Public Prosecutions* [1937] AC 576.

*Attorney-General for Northern Ireland's Reference (No. 1 of 1975)* [1977] AC 105.

*Attorney-General for Northern Ireland v Gallagher* [1963] AC 349.

*Attorney-General's Reference (No. 2 of 1992)* [1994] QB 91.

*Attorney-General's Reference (No. 1 of 2002)* [2002] EWCA Crim 2392.

*Attorney-General v Scotcher* [2005] UKHL 36.

*Atwal v Massey* (1972) 56 Cr App R 6.

*Boggeln v Williams* (1978) 67 Cr App R 50.

*Bolam v Friern Hospital Management Committee* [1957] 2 All ER 118.

*Bowman v Blyth* (1856) 7 E&B 26.

*Bratty v Attorney-General for Northern Ireland* [1963] AC 386.

*Buckoke v GLC* [1971] Ch 655.

*Caparo Industries plc v Dickman* [1990] 2 AC 605.

*Chandler v Director of Public Prosecution* [1964] AC 763.

*Chief Constable of Avon and Somerset v Shimmen* (1987) 84 Cr App R 7.

*Churchill v Walton* [1967] 2 AC 224.

*Commissioner of Police of the Metropolis v Caldwell* [1982] AC 341.

*Cunliffe v Goodman* [1950] 2 KB 237.

*Daniel M'Naghten's Case* (1843) 8 ER 718.

*Davis Contractors Ltd. v Fareham Urban District Council* [1956] AC 696.

*Devlin v Armstrong* [1971] NILR 13.

*Director of Public Prosecutions for Northern Ireland v Lynch* [1975] AC 653.

*Director of Public Prosecutions v Armstrong-Braun* [1999] Crim LR 416.

*Director of Public Prosecutions v B* [2000] 2 AC 428.

*Director of Public Prosecutions v Beard* [1920] AC 479.

*Director of Public Prosecutions v Little* [1992] QB 645.

*Director of Public Prosecutions v Majewski* [1977] AC 443.

*Director of Public Prosecutions v Morgan* [1976] AC 182.

*Director of Public Prosecutions v Pipe* [2012] EWHC 1821.

*Director of Public Prosecutions v Smith* [1961] AC 290.

*Director of Public Prosecutions v Smith* [2006] EWHC 94.

*Donoghue v Stevenson* [1932] AC 562.

*Elliott v C* [1983] 1 WLR 939.

*Fagan v Metropolitan Police Commissioner* [1969] 1 QB 439.

*Gillick v West Norfolk and Wisbech AHA* [1986] AC 112.

*Gregson v Gilbert* (1783) 3 Doug. KB 232.

*Hall v Brooklands Auto Racing Club* [1933] 1 KB 205.

*Haughton v Smith* [1975] AC 476.

*Healthcare at Home Ltd. v The Common Services Agency* [2014] UKSC 49.

*Herrington v British Railways Board* [1971] 2 QB 107.

*Hill v Baxter* [1958] 1 QB 277.

*Hills v Ellis* [1983] QB 680.

*Hussien v Chang Fook Kam* [1970] AC 942.

*Hyam v Director of Public Prosecutions* [1975] AC 55.

*Iceland Frozen Foods Ltd. v Jones* [1983] ICR 17.

*Inner South London Coroner, ex parte Douglas-Williams* [1999] 1 All ER 344.

*Ivey v Genting Casinos (UK) Ltd.* [2016] EWCA Civ 1093.

*Ivey v Genting Casinos (UK) Ltd. t/a Crockfords* [2017] UKSC 67.

*Johnson v Youden* [1951] 1 KB 544.

*Lee v Taylor and Gill* (1912) 77 JP 66.

*Loake v Director of Public Prosecutions* [2018] QB 998.

*London Borough of Southwark v Williams* [1971] 2 All ER 175.

*Maynard v West Midlands Regional Health Authority* [1985] 1 All ER 635.

*M. J. J. (a minor) v Cooper* (unreported) (CO/1551/84, 2 July 1987).

*Moore v Hussey* (1609) Hob 96.

*Mouse's Case* (1620) 12 Co Rep 63.

*Paul v Ministry of Posts and Telecommunications* [1973] RTR 245.

*Perrett v Collins* [1998] 2 Lloyds Rep 255.

*R (on the application of Countryside Alliance) v Attorney-General* [2008] AC 719.

*R (on the application of Nicklinson) v Ministry of Justice* [2014] UKSC 38.

*R (on the application of Pretty) v Director of Public Prosecutions* [2002] 1 AC 800.

*R (on the application of the RSPCA) v C* [2006] EWHC 1069.

*Re A (conjoined twins)* [2001] Fam 147.

*Re Continental Assurance Co of London plc* [2007] 2 BCLC 287.

*Re F (Mental Patient: Sterilisation)* [1990] 2 AC 1.

*Re T (Minors) (Custody: Religious Upbringing)* (1981) 2 FLR 239.

*Royal Brunei Airlines Sdn Bhd v Tan* [1996] 2 AC 378.

*R v Acott* [1997] 2 Cr App R 94.

*R v Adomako* [1995] 1 AC 171.

*R v Ali* [1989] Crim LR 736.

*R v Allen* [1988] Crim LR 698.

*R v B* [2013] EWCA Crim 3.

*R v Bailey* [1983] EWCA Crim 2.

*R v Bannister* [2009] EWCA Crim 1571.

*R v Bateman* (1927) 19 Cr App R 8.

*R v Bates* [1952] 2 All ER 842.

*R v Beckford* [1988] AC 130.

*R v Belfon* (1976) 63 Cr App R 59.

*R v Bell* [1984] 3 All ER 842.

*R v Bello* (1978) 67 Cr App R 288.

*R v Bennett* [1995] Crim LR 877.

*R v Bingham* [1991] Crim LR 43.

*R v Bourne* [1939] 1 KB 687.

*R v Bournewood Community and Mental Health NHS Trust* [1998] 3 All ER 289.

*R v Bowen* (1996) 2 Cr App R 157.

*R v Brady* [2006] EWCA Crim 2413.

*R v Brandford* [2016] EWCA Crim 1794.

*R v Briggs* [1977] 1 WLR 605.

*R v Brown* [2011] EWCA Crim 2796.

*R v Burgess* [1991] 2 QB 92.

*R v Burns* (1974) 58 Cr App R 364.

*R v Buswell* [1972] 1 WLR 64.

*R v Byrne* [1960] 2 QB 396.

*R v Caldwell* (1980) 71 Cr App R 237.

*R v Chisam* (1963) 47 Cr App R 130.

*R v Church* [1966] 1 QB 59.

*R v Clarke* [1972] 1 All ER 219.

*R v Coley, McGhee and Harris* [2013] EWCA Crim 223.

*R v Colohan* [2001] EWCA Crim 1251.

*R v Coney* (1882) 8 QBD 534.

*R v Conway* [1989] QB 290.

*R v Cox* (1992) 12 BMLR 38.

*R v Cugullere* (1961) 45 Cr App R 108.

*R v Cunningham* [1957] 2 QB 396.

*R v Da Silva* [2006] EWCA Crim 1654.

*R v Dawes* [2013] EWCA Crim 322.

*R v Dietschmann* [2003] 1 AC 1209.

*R v Dodman* [1998] 2 Cr App R 338.

*R v Dudley and Stephens* (1884) 14 QBD 273.

*R v Eatch* [1980] Crim LR 650.

*R v Faulkner* (1877) 13 Cox CC 550.

*R v Feely* [1973] QB 530.

*R v Fennell* [1971] 3 All ER 215.

*R v Fenton* (1975) 61 Cr App R 261.

*R v Foster* (1825) 1 Lewin 187.

*R v G* [2003] UKHL 50.

*R v GAC* [2013] EWCA Crim 1472.

*R v Ghosh* [1982] QB 1053.

*R v Gilks* (1972) 56 Cr App R 734.

*R v Gittins* [1984] QB 698.

*R v Golds* [2016] UKSC 61.

*R v Goodwin* [2018] EWCA Crim 2287.

*R v Graham* (1982) 74 Cr App R 235.

*R v Griffiths* (1974) 60 Cr App R 14.

*R v Gurpinar* [2015] EWCA Crim 178.

*R v Hall* (1985) 81 Cr App R 260.

*R v Hammond* [2013] EWCA Crim 2709.

*R v Hancock and Shankland* [1986] AC 455.

*R v Hardie* (1985) 80 Cr App R 157.

*R v Hasan* [2005] UKHL 22.

*R v Hayes* [2015] EWCA Crim 1944.

*R v Hennessy* [1989] 1 WLR 287.

*R v Hinks* [2001] 2 AC 241.

*R v Howe* [1987] AC 417.

*R v Howell* (1938) 27 Cr App R 5.

*R v Hudson* [1966] 1 QB 448.

*R v Hudson and Taylor* [1971] 2 QB 202.

*R v Hurst* (1995) 1 Cr App R 82.

*R v Ireland* [1998] AC 147.

*R v Jakeman* (1983) 76 Cr. App. R. 223.

*R v Jewell* [2014] EWCA Crim 414.

*R v Johnson* [2007] EWCA Crim 1978.

*R v K* [2001] UKHL 41.

*R v Keane* [2010] EWCA Crim 2514.

*R v Kemp* [1957] 1 QB 399.

*R v Kennedy* [2007] UKHL 38.

*R v Kimber* [1981] 3 All ER 84.

*R v Kingston* [1995] 2 AC 355.

*R v Kitson* (1955) 39 Cr App R 66.

*R v Kuddus* [2019] EWCA Crim 837.

*R v Lamely* (1911) 22 Cox CC 635.

*R v Lane and Letts* [2018] UKSC 36.

*R v Lawrence* [1982] AC 510.

*R v Le Brun* [1991] 4 All ER 673.

*R v Lloyd* [1967] 1 QB 175.

*R v Marison* [1997] RTR 457.

*R v Martin* [1881] 8 QB 54.

*R v Martin* (1989) 88 Cr App R 343.

*R v Martindale* (1987) 84 Cr App R 31.

*R v McCalla* (1988) 87 Cr App R 372.

*R v Miller* [1983] 2 WLR 539.

*R v M'Naughten* (1843) 8 ER 718.

*R v Mohan* [1976] QB 1.

*R v Moloney* [1985] 1 AC 905.

*R v Montila* [2004] UKHL 50.

*R v Murphy* [1980] QB 434.

*R v Nedrick* [1986] 1 WLR 1025.

*R v Oatridge* (1992) Cr App R 367.

*R v Olugboja* [1982] QB 320.

*R v Owino* (1996) 2 Cr App R 128.

*R v Palmer* [1971] AC 814.

*R v Parker* [1977] 1 WLR 600.

*R v Pattni* (unreported) [2001] *Criminal Law Review* 570.

*R v Pearman* (1984) 80 Cr App R 259.

*R v Pearson* (1835) 2 Lewin 144.

*R v Pembliton* (1874) LR 2 CCR 119.

*R v Philpot* (1912) 7 Cr App R 140.

*R v Pleydell* [2005] EWCA Crim 1447.

*R v Price* (1884) 12 QBD 247.

*R v Price and Bell* [2014] EWCA Crim 229.

*R v Prince* (1875) LR 2 CCR 154.

*R v Quayle* [2005] EWCA Crim 1415.

*R v Quick* [1973] QB 910.

*R v R* [1991] UKHL 12.

*R v Ray* [2017] EWCA Crim 1391.

*R v Reid* [1992] 1 WLR 793.

*R v Rejmanski* [2017] EWCA Crim 2061.

*R v Richardson and Irwin* (1999) 1 Cr App 392.

*R v Rimmington* [2005] UKHL 63.

*R v Rose* (1884) 15 Cox 540.

*R v R (Stephen Malcolm)* (1984) 79 Cr App R 334.

*R v Russell* (1985) 81 Cr App R 315.

*R v Saik* [2006] UKHL 18.

*R v Savage and Parmenter* [1991] 1 AC 699.

*R v Seymour* [1983] 2 AC 493.

*R v Simcox* [1964] Crim LR 402.

*R v Smith* (1837) 8 Car & P 158.

*R v Stephenson* [1979] 1 QB 695.

*R v Stewart* [2009] EWCA Crim 593.

*R v Sullivan* [1984] 1 AC 156.

*R v Tandy* (1988) 87 Cr App R 45.

*R v Thabo-Meli* [1954] 1 WLR 228.

*R v Tolson* (1889) 23 QBD 168.

*R v Venna* [1976] QB 421.

*R v Walker and Hayles* (1990) 90 Cr App R 226.

*R v Ward* [1956] 1 QB 351.

*R v Welch* [1875] 1 QB 23.

*R v West Berkshire Health Authority* [1989] 2 AC 1.

*R v Weston* (1879) 14 Cox CC 346.

*R v Wilcocks* [2016] EWCA Crim 2043.

*R v Williams* [1953] 1 All ER 1068.

*R v Williams (Gladstone)* (1984) 78 Cr App R 276.

*R v Williamson* (1807) 172 ER 579.

*R v Windle* [1952] 2 QB 826.

*R v Wood* [2008] EWCA Crim 1305.

*R v Woods* (1982) 74 Cr App R 312.

*R v Woodward* [1995] 2 Cr App R 388.

*R v Woollin* (1997) 1 Cr App R 97.

*R v Woollin* [1999] 1 AC 82.

*R v Wright* [2000] Crim LR 510.

*R v Yip Chiu-Cheung* [1995] 1 AC 111.

*Scott v Metropolitan Police Commissioner* [1975] AC 819.

*Shawinigan Ltd. v Vokins & Co. Ltd.* [1961] 1 WLR 1206.

*Simpson v Peat* [1952] 2 QB 24.

*Smith v Littlewoods* [1987] UKHL 18.

*Stovin v Wise* [1996] AC 923.

*Sweet v Parsley* [1970] AC 132.

*Taylor's Central Garages (Exeter) v Roper* [1951] 2 TLR 284.

*Watmore v Jenkins* [1962] 2 QB 572.

*Westminster City Council v Croyalgrange Ltd.* (1986) 83 Cr App R 155.

*Woods v Richards* (1977) 65 Cr App R 300.

*Yuen Kun Yeu v Attorney-General of Hong Kong* [1988] AC 175.

*

*Morissette v United States* (1952) 342 US 246, 341 (US).

*R v Codère* (1916) 12 Cr App R 21 (Canada).

*R v Chaulk* [1990] 3 SCR 1303 (Canada).

*R v George* (1960) 128 Can CC 289 (Canada).

*R v Perka* [1984] 2 SCR 232 (Canada).

*R v Porter* [1936] 55 CLR 182 (Australia).

*R v Ruzic* [2001] SCR 687 (Canada).

*R v Ryan* [2013] SCC 3 (Canada).

*R v Weise* [1969] VR 953 (Australia).

*Stapleton v R* (1953) 86 CLR 358 (Australia).

## ***Legislation***

14 Edw. III Stat. 1 c.4 (1340).

Animals (Scientific Procedures) Act 1986.

Armed Forces Act 2006.

Companies Act 2006.

Coroners and Justice Act 2009.

Crime and Courts Act 2013.

Crime and Disorder Act 1998.

Criminal Damage Act 1971.

Criminal Finances Act 2017.

Criminal Justice Act 1967.

Criminal Justice Act 1988.

Criminal Justice Act 2003.

Criminal Justice and Immigration Act 2008.

Criminal Law Act 1967.

Criminal Law Act 1977.

Criminal Procedure (Insanity) Act 1964.

Domestic Violence, Crime and Victims Act 2004.

Equality Act 2010.

Firearms Act 1968.

Fraud Act 2006.

Gambling Act 2005.

Homicide Act 1957.

Hunting Act 2004.

Immigration Act 1971.

Larceny Act 1916.

Legal Aid, Sentencing and Punishment of Offenders Act 2012.

Local Government (Miscellaneous Provisions) Act 1982.

Misuse of Drugs Act 1971.

Motor Car Act 1903.

Offences Against the Person Act 1861.

Police and Criminal Evidence Act 1984.

Prevention of Crime Act 1953.

Proceeds of Crime Act 2002.

Protection from Harassment Act 1997.

Protection of Animals Act 1911.

Psychoactive Substances Act 2016.

Public Order Act 1986.

Road Traffic Act 1930.

Road Traffic Act 1972.

Road Traffic Act 1988.

Serious Crime Act 2007.

Sexual Offences Act 2003.

Suicide Act 1961.

Terrorism Act 2000.

Theft Act 1968.

Trial of Lunatics Act 1883.

Unsolicited Goods and Services Act 1971.

Vagrancy Act 1824.

*

Assisted Dying Bill 2014.

Assisted Dying Bill (No. 2) 2015-16.

Assisted Dying Bill 2016-17.

Assisted Dying Bill 2021.

Patient (Assisted Dying) Bill 2002.

Terminally Ill Bill 2004.

*

Explanatory Notes to the Fraud Act 2006.