Universitat de Girona

LUCK AND THE CONTROL THEORY OF KNOWLEDGE

Fernando BRONCANO-BERROCAL

Dipòsit legal: Gi. 247-2014
http://hdl.handle.net/10803/129460

Universitat de Girona

LUCK AND THE CONTROL THEORY OF KNOWLEDGE

DOCTORAL THESIS

FERNANDO BRONCANO-BERROCAL

2013

Fernando Broncano-Berrocal

Luck and the Control Theory of Knowledge

Doctoral Thesis, 2013

PROGRAM:
Ciències Humanes i de la Cultura

SUPERVISORS:
Joan Pagès (Universitat de Girona)
Manuel Pérez Otero (Universitat de Barcelona)

DEPARTAMENT:
Departament de Filosofia

FACULTY:
Facultat de Lletres

UNIVERSITY:
Universitat de Girona

Memòria presentada per optar al títol de doctor per la Universitat de Girona

A mis padres

I approach deep problems such as I do cold baths: fast in, fast out. That this is no way to get to the depths, to get deep enough, is the superstition of those who fear water, the enemies of cold water; they speak without experience. Oh, the great cold makes one fast!

Nietzsche ("la gaya scienza")

ABSTRACT

This thesis presents a diagnosis of the problem of luck in epistemology and an analysis of the concept of knowledge. Part I gives an account of the ordinary concept of luck. Part II gives an account of the philosophical notion of epistemic luck and develops an original account of the concept of knowledge: the control theory of knowledge.

RESUM

Aquesta tesi presenta un diagnòstic del problema de la sort en epistemologia i una anàlisi del concepte de coneixement. La primera part ofereix una teoria del concepte ordinari de sort. La segona part ofereix una teoria de la noció filosòfica de sort epistèmica i desenvolupa una teoria original del concepte de coneixement: la teoria del control.

RESUMEN

Esta tesis presenta un diagnóstico del problema de la suerte en epistemología y un análisis del concepto de conocimiento. La primera parte ofrece una teoría del concepto ordinario de suerte. La segunda parte ofrece una teoría de la noción filosófica de suerte epistémica y desarrolla una teoría original del concepto de conocimiento: la teoría del control.

PUBLICATIONS

Some ideas appear in:

Broncano-Berrocal, F. (2013). Lies and deception: A failed reconciliation. *Logos & Episteme. An International Journal of Epistemology*, 4, 227–230.

Broncano-Berrocal, F. (*forthcoming*). Is safety in danger? *Philosophia*.

DECLARATION

El Dr. Joan Pagès Martínez de la Universitat de Girona i el Dr. Manuel Pérez Otero de la Universitat de Barcelona

DECLAREM:

Que el treball titulat *Luck and the Control Theory of Knowledge*, que presenta Fernando Broncano-Berrocal per a l'obtenció del títol de doctor, ha estat realitzat sota la meva direcció i que compleix els requisits per poder optar a Menció Internacional.

I, perquè així consti i tingui els efectes oportuns, signo aquest document.

Joan Pagès	Manuel Pérez Otero

Girona,

ACKNOWLEDGMENTS

Writing this dissertation has been like making an immense jigsaw puzzle, and several people have been of great help to put all the pieces together. I am very grateful to my supervisors Joan Pagès and Manuel Pérez Otero, not only because their thorough comments on a considerable number of considerable long drafts have made me avoid more than a considerable number of mistakes, but also because their helpful and constructive criticism has contributed to my learning of how to write philosophy. When one does a M.Phil and a PhD at LOGOS, it is very difficult not to learn something new every day, and many things I have learned about many fascinating topics have helped me to give shape to the project, but from Logosians, both seniors and juniors, I have mainly learned to practice philosophy cultivating clarity and accuracy of thought, and I am very grateful to them.

When I have been stuck with some ideas, presenting them at the Universities of Barcelona and Girona and at several other places in Europe has been of great help. I am particularly grateful to the kind members of the COGITO Research Centre in Philosophy at the University of Bologna who gave me the opportunity to present and discuss a lot of material with them in a very friendly manner.

Outside philosophy, I've shared home in Barcelona with a snail, a tiger and a little tree. They know who they are, they are great friends and it's been a lot of fun. I hope they will finish their PhD's soon, or that they have already forgotten about the PhD's that they already finished, because, in the end, things that matter most do not need a certificate.

I can now understand my sister Alicia and how hard should have been to spend so many hours wondering how light is a neutrino or, in my case, what is this thing called knowledge. I am thankful to her for being my older sister and for paving the way for me, with everything that implies.

This thesis is dedicated to my parents because they have always been there for us. My mother is a constant source of care, love and support and she is the best mother a son could ever have. Endless conversations with my father have been the best help I could imagine during the writing process and he should take credit for many of the ideas of the dissertation. Without the support of my parents, I would not have been able to complete this project.

Fede has been there with me during this long journey. She is the most extraordinary person I've ever met. And I love her so much.

PREFACE

In 1963, Edmund Gettier described in a three-page paper a couple of ordinary scenarios in which someone has good justification to believe something true, believes it, but does not know it. Two surprisingly mundane cases challenged a philosophical assumption a couple of thousand of years old: that three criteria (truth, belief and justification) are sufficient to state the nature of propositional knowledge. Needless to say, the amazing brevity of the paper was inversely proportional to the quantity of ink used to analyze it. Epistemologists soon agreed that the subjects described by Gettier came to believe the truth by accident, that despite having excellent justification to believe what they believed it was just a coincidence that they came to believe something true. The intuition that everyone had was that one cannot have knowledge if one comes to believe the truth by luck and the problem that everyone noticed was that no answer to the question of what is knowledge could ever be satisfactory if that intuition was overlooked. This is how the problem of luck in epistemology came to existence.

We can compare the problem of luck to a disease. In 1963, the disease was exotic and its reach was barely known. When a disease appears for the first time, doctors run tests on their patients, but since little is known about what causes the disease, they have no other option but to apply treatment after treatment until their patients heal. In the same way, the history of the problem of luck in epistemology has been a trial-and-error process. In the early years, epistemologists knew little about the problem so that, after some preliminary diagnoses, they attempted to heal the patient by applying epistemic condition after epistemic condition: the notion of justification was thus interpreted in very specific ways and the triad <truth, belief and justification> was supplemented with new conditions (e.g., no false lemmas solutions, defeasibility analyses, causal accounts or simple reliabilism).

However, the pervasive problem of luck manifested in different new guises and for each new analysis of knowledge that appeared, a new counterexample emerged, and yet the underlying intuition remained the same: one cannot have knowledge if one comes to believe the truth by luck. So-called Gettier-style cases turned out into a sort of test that any successful analysis of knowledge should pass. As a result, the literature became populated with complex analyses of knowledge, creative counterexamples and a rich family of epistemic concepts.[1]

1 See Robert K. Shope's 1983 remarkable book for an overall picture of this escalation of analyses and corresponding counterexamples in the early post-Gettier years.

As the years went by, new treatments were developed and epistemologists thought to have found epistemic conditions immune to the disease (e.g., modal and virtue-theoretic conditions). However, those epistemic conditions led to new problems and some epistemologists, horrified by the secondary effects of these new treatments and by a recent history of disagreements and controversies, decided to try an alternative medicine: the very analyzability of the concept of knowledge was cast into doubt.

Fortunately, there is no need to apply such radical treatments. As it often happens in medicine, where new research in other fields allows to understand the causes and behavior of a disease, the problem of luck in epistemology has been illuminated by recent research on the general notion of luck. And there is reason to be optimistic: we now count with several definitions of luck that help us understand under what conditions a belief is true by luck.

This dissertation pursues an ambitious research project: to diagnose the problem of luck in epistemology and to offer an analysis of the concept of knowledge. The structure of the dissertation reflects the three stages that, according to Duncan Pritchard (2009a: 33), are essential to the development of an anti-luck epistemology:[2]

1. Give an account of luck (chapters 1, 2 and 3).
2. Specify the sense in which knowledge is incompatible with luck (chapter 4).
3. Show what conditions must be satisfied in order to block the kind of luck incompatible with knowledge and incorporate them within a theory of knowledge (chapters 5, 6, 7, 8 and 9).

The following is an overview of its contents:

Chapter 1 analyzes the chance condition for luck (the condition that an event is lucky for an agent only if the event is chancy). Several possible ways of understanding 'chancy' are explored: in terms of 1) accidentality, 2) coincidence, 3) indeterminacy, 4) subjective probability, 5) epistemic probability, 6) objective probability and 7) modality. The notion of risk is analyzed and two senses of risk are distinguished: agent-focused and belief-focused risk. It is argued that luck must be understood in terms of both senses of risk (luck is a specific form of risk). Agent-focused risk is defined in terms of lack of control. Two conceptions of event-focused risk are considered: in terms of objective probability and in modal terms. It is argued that the kind of event-focused risk that serves to define luck is modal, not probabilistic. Chancy events are thus defined as events that are risky in the modal sense. Finally, it is argued that there is a set of events that are not risky in this modal sense but yet involve some 'luckiness'. They

2 The description of the stages is from Carter (2010: 518).

receive the label of 'fortunate' events. A crucial distinction between luck and fortune is made.

Chapter 2 analyzes the significance condition for luck (the condition that an event is lucky for an agent only if the event is significant for the agent). It is argued that the condition is necessary for luck. Several formulations of the condition are discussed. One of them is defended. Definitions of luck and fortune are provided.

Chapter 3 analyzes the lack of control condition for luck (the condition that an event is lucky for an agent only if the agent lacks control over the event). A lack of control condition in terms of lack of choice is discussed and rejected. A general account of the notion of control is offered. Two types of control are distinguished: effective and tracking control. Several objections to conceiving luck in terms of lack of control are discussed and rejected.

Chapter 4 analyzes the notion of epistemic luck. The first part of the chapter discusses Unger/Pritchard's taxonomy of epistemic luck. Special attention is paid to the notions of veritic and reflective epistemic luck. In the second part of the chapter, epistemic luck is explained in terms of epistemic risk. Two types of epistemic risk are distinguished: belief-focused risk and agent-focused epistemic risk. The agent-focused sense of epistemic risk is defined in terms of lack of epistemic control. The belief-focused sense of epistemic risk is defined in terms of modal fragility. Three types of belief-focused risk are distinguished. A distinction is made between veritic epistemic luck and veritic epistemic fortune. Paradigmatic cases of the literature are analyzed in the light of that distinction and an overarching hypothesis about knowledge is advanced: nearly all (if not all) cases of true belief that is not knowledge are cases either of veritic luck or of veritic fortune. It is explained that the problem of luck in epistemology is much more pervasive than it was initially thought. It is explained that veritic luck and veritic fortune arise when an agent lacks epistemic control and the core working hypothesis of the rest of the dissertation is stated: knowledge requires epistemic control.

Chapter 5 analyzes the modal approach to knowledge with the purpose of translating the notion of tracking control in epistemic terms. Nozick's tracking metaphor of knowledge as tracking the truth is presented and developed. It is argued that the notion of epistemic control has a modal side: an agent controls her cognitive performance only if her beliefs would succeed or not fail in a certain range of possibilities. Three epistemic conditions that delimit (force) the set of possibilities over which an inquiring agent has to succeed or not fail are analyzed: sensitivity, receptivity and safety. The pros and cons of each condition are discussed. It is concluded that safety is the best way of specifying the modal side of epistemic control: although it fails to exclude veritic fortune, it is the anti-luck condition per excellence.

Chapter 6 analyzes the achievement account of knowledge (the view that knowledge is a cognitive achievement). The chapter starts with the basics: a presentation of the virtue-theoretic approach to epistemology, of Ernest Sosa's performance-assessment model and of the idea that knowledge is a cognitive achievement. The standard notion of achievement (achievement as success because of ability) is analyzed. An alternative account in terms of control is advanced (achievement as success over which the agent has control). Two objections by Duncan Pritchard against the standard notion of achievement and against its corresponding standard achievement account of knowledge are considered (the problem of lucky achievements and the problem of easy achievements). It is argued that the control conception of achievement can steer clear of both problems. A non-standard version of the achievement account of knowledge is proposed: the so-called control theory of knowledge, which entails (but is broader than) the standard achievement account of knowledge.

Chapter 7 argues that the kind of virtue-theoretic conditions for knowledge offered by proponents of the standard achievement account of knowledge are especially well suited to account for an epistemic analogue of the notion of effective control. Several of those conditions are presented as necessary for epistemic control: an agent controls her cognitive performance only if 1) it is successful, 2) arises out of cognitive competence that is 3) well integrated with the rest of the agent's cognitive competences and 4) it is successful because of competent. Several ways of cashing out the 'because of' relation are discussed. It is argued that an explanatory reading of 'because of' leads to incorrect judgments of certain test cases. It is argued that 'because of' must be read in terms of manifestation of competence (aptness). Two views on aptness are distinguished: the appropriateness view and the situational view. It is argued that only the latter is correct. In addition, it is explained the sense in which the notion of epistemic control offered does not entail doxastic voluntarism. Finally, two ways of explaining lottery cases in terms of lack of epistemic control are discussed.

Chapter 8 presents the hypothesis that epistemic control arises when the safety of a belief is because of epistemic competence. Several accounts of knowledge that combine safety with a virtue-theoretic condition are discussed. Christoph Kelp's safe-apt view (knowledge as safe, apt belief) is analyzed. It is argued that beliefs might be safe, apt, and yet not knowledge. Duncan Pritchard's anti-luck virtue epistemology is analyzed. The theory is read in two ways: 1) knowledge as the conjunction of safety and the condition that the agent's cognitive success must be partially creditable to cognitive ability and 2) knowledge as safe belief whose safety is partially creditable to the agent's cognitive abilities. It is argued that, in either of the two readings, the theory is not sufficient for knowledge. In addition, a problem concerning the direction of fit of known beliefs is presented. It is argued that only the situational view of aptness is able to solve it. An account of inferential knowledge in terms of aptness is sketched. Finally, John Turri's proposal of knowledge as ample belief (belief whose safety manifests competence) is analyzed. It is argued that the account lacks specificity and that its motivation is not adequate.

Chapter 9 provides an analysis of the concept of knowledge. In the first part of the chapter, the safety condition is defended from a recent counterexample by Tomas Bogardus to its necessity for knowledge. It is argued that the case just corroborates the well-known requirement that modal conditions must be relativized to methods of belief formation. An externalist principle of method individuation and a corresponding version of the safety condition is proposed. It is explained that Bogardus's case is not a counterexample to the necessity of safety but to the necessity of a condition called super-safety. In the second part of the chapter, Sosa's performance-assessment model is extended with a normative property called security. A secure performance is an apt performance that would not easily be inapt were the agent to perform in the same way as she actually does. A distinction is made between security and another normative property called super-security. Epistemic control is defined in terms of secure cognitive performance. The notion of cognitive achievement is defined in terms of epistemic control. Knowledge is defined as a cognitive achievement. Knowledge is defined as secure belief.

CONTENTS

Part I

LUCK AND FORTUNE

PRELIMINARY REMARKS

Philosophers who have theorized about the concept of luck have characterized it with three types of conditions: lack of control conditions, chance conditions and significance conditions. *Lack of control conditions* capture the idea that an event is lucky for an agent only if the agent lacks control over the event. The core idea of *chance conditions* is that lucky events are chancy events. Finally, *significance conditions* introduce the idea that an event, even if chancy or beyond the agent's control, cannot be regarded as lucky if it is not significant for her.

In the next three chapters, I will present a detailed analysis of the different versions of these three types of conditions and of the different accounts of luck that result from combining two or three of them.[3] Many detailed points will be covered and made, but all will be oriented towards a very simple idea: luck should be defined in terms of risk (in a sense that will be specified).

Before entering into details, it is convenient to highlight some of the general features of luck. First, the notion of luck applies to *events* and *states* (for discussion purposes, I will talk about luck as a phenomenon that applies to events). Second, an event cannot be conceived as lucky if it is not conceived as lucky *for an individual*. In other words, luck is always relativized to individuals. Third, there is a sense in which a group of individuals can be said to be lucky, as when we say that a group of rock-climbers is lucky to have survived a fortuitous rockfall. But as E. J. Coffman (2007: 386) remarks, there seems to be no reason why *group luck* cannot be reduced to or explained in terms of *individual luck.*

Fourth, the kind of beings to which we ascribe luck are *beings with interests* (Ballantyne 2012). A human or a dog are lucky to have survived a fortuitous rockfall; a stick of wood or a car are not. Still, at least in some contexts, it seems correct to ascribe luck to an object without interests, as when we say that our beloved car is lucky not to have been damaged by a fortuitous rockfall. However, this kind of assertions are felicitous insofar as they are parasitic on our interests. We would not say that a stick of wood is lucky not to have been destroyed by a rockfall if its existence bore absolutely no significance for us, and if we would say that, we would only say it figuratively.

Fifth, luck is a gradual notion. In ordinary language, it is common to ascribe different *degrees of luck* to different events. Interestingly, the three conditions (significance, chance and lack of control) may in prin-

3 Significance conditions are the only conditions considered necessary for luck by all commentators and they are combined with some version of the chance condition, of the lack of control condition or of both of them to define the concept of luck. In my view, the three types of conditions are needed to define it.

ciple contribute to the explanation of the gradualness of luck. For instance, consider lottery *A* and lottery *B*. By stipulation, the chance of winning lottery *A* is the same as the chance of winning lottery *B*, but *A*'s prize is surviving while *B*'s prize is a big amount of money. Intuitively, winning lottery *A* is luckier than winning lottery *B*. One is tempted to explain this intuition by appealing to the fact that for most of us, humans, it is more important to be alive than to be rich. This sort of explanations are entailed by versions of the significance condition.

On the other hand, imagine that someone buys a lottery ticket and wins. Now imagine that in other occasion other person buys one thousand tickets of the same lottery and wins. Intuitively, the former person is luckier than the latter. One might explain this divergence of degree by elaborating the idea that the 'chancier' an event is (in a sense that is not necessarily probabilistic), the luckier the event is. In principle, some gradual version of the chance condition should be able to accommodate this idea.

The same kind of considerations apply to the lack of control condition. If one thinks that an agent's lack control over an event is necessary for the event to be lucky for the agent, one will be willing to explain that a goal from the midfield by an amateur player is luckier than a goal from the same position by, say, Lionel Messi, because Messi, who is the best soccer player in the planet, is able to exert a greater degree of control over the ball's trajectory than the amateur. In sum, the explanation of why luck is a gradual notion depends on which conditions one takes to be necessary for luck.

Sixth, while in ordinary language we typically apply the term 'lucky' to events that are perceived as beneficial, a common stipulation in philosophical contexts is that the term 'lucky' is to be used to refer to events that instantiate *good luck* (e.g., surviving a fortuitous rockfall), but also to refer to events that instantiate *bad luck* (e.g., losing a large sum of money in a safe bet).

Seventh, a final caveat: luck is a *vague concept*. Not all instances of luck will be as clear-cut as a lottery win. We can expect some cases in which our intuitions conflict. For example, consider goals from the corner kick in professional soccer matches. Are they by luck? It depends on the case but, in general, intuitions tend to be divided. Consider the following example given by Duncan Pritchard (2005: 143): suppose that *S* drops her wallet, keeps walking and after five minutes returns and finds the wallet in the place where she dropped it. It is by luck that *S* has found her wallet? The answer is not clear. Accordingly, we should not expect of our analysis of luck to remove this vagueness. On the contrary, it should predict borderline cases.

1 THE CHANCE CONDITION

Consider the event of winning a fair lottery, a paradigmatic instance of luck. *S* buys a lottery ticket and wins. Since the lottery has been fair, it is by chance that *S* has won. It seems that it is in virtue of its being by chance that *S*'s winning is by luck. Imagine now that *S* rigs the lottery and wins. We would not say that her winning is lucky. In fact, when one rigs a lottery in one's favor, it might be found intuitive to appeal to the fact that the lottery's outcome is no longer by chance to explain that it is not by luck that one wins. Thus, there seems to be a necessary link between luck and chance. Roughly, an event is lucky only if it is by chance. Let us call this condition the *chance condition* for luck.[1]

It is important to keep in mind that the condition is necessary but not sufficient for luck. All lucky events are by chance, but it would be an error to hold that all events that are by chance are lucky. A rockfall that happens by chance is not lucky if it does not have positive or negative effects on anyone. In this sense, the chance condition needs to be supplemented at least with a condition on the significance of the relevant event (i.e., with a significance condition).

Depending on what is thought to determine the chance of an event occurring, the chance condition might be formulated in one way or another. There are five main ways of specifying the condition: 1) In terms of *accidentality*: the idea is that whatever makes an event accidental is what makes it chancy. 2) In terms of *indeterminacy*: the idea here is rather that chancy events are events that were not determined to occur prior to their occurrence. 3) In terms of *subjective probability*: under this interpretation, chance is cashed out in terms of what is expected to occur by some agent. 4) In terms of *epistemic probability*: this view states that chancy events are events that are not likely to occur given the available evidence about them. 5) In terms of *risk*: the idea is to conceive chancy events as events at risk of occurring (or of not occurring), where risk is defined either in terms of *objective probability* or in *modal* terms (lucky events in this sense would be events that obtain but that were at risk of not obtaining).

In the following sections, I will present arguments against 1-4. I will also argue that the chance condition should not be formulated

1 Unless otherwise indicated, with the term 'chance' I do not mean physical or objective probabilities (chances), neither when I use it to name the condition for luck addressed in this chapter, nor when I speak of the chance of an event occurring or of chancy events. I use the term in a pre-theoretical way so that it can be subsequently explained in terms of more technical notions being objective probabilities one of them.

in terms of probabilistic risk. My aim is to argue for a chance condition in terms of modal risk. While discussing accidentality, I will also present some arguments against the possibility of defining lucky events (not just chancy events) as accidents, and I will also explore the relation between luck and coincidences and the possibility of conceiving luck as an epistemic notion.

1.1 ACCIDENTS AND ACCIDENTALITY

The first way of understanding chance is in terms of accidentality. According to this view, chancy events are events whose occurrence is *accidental*:

- ACCIDENTAL CHANCE: *E* is lucky for *S* at *t* only if it is accidental that *E* occurs at *t*.[2]

The view has some initial plausibility. On the one hand, it captures the intuition that finding a coin on the sidewalk as one is going to work is a lucky event (because one finds it by accident). On the other hand, the view might be used to explain Gettier-style cases (albeit in a not very informative way). By way of illustration, consider the following two Gettier-style cases: 1) *S* forms the belief that there is a sheep in the field on the basis of misleading evidence (e.g., *S* looks at a dog that looks like a sheep); unbeknownst to *S*, there is a sheep hidden behind one of the trees of the field.[3] Diagnosis: *S* does not know that there is a sheep in the field because it is accidental that *S* believes that proposition while, unbeknownst to her, there is a sheep behind a tree. 2) *S* looks at an egg box that she has just bought and forms the belief that one particular egg is fresh (the rest are boiled eggs). Diagnosis: *S* does not know that the egg is fresh because it is accidental that she looks at the only fresh egg in the box.

Note that the view that an event occurs by chance if and only if it occurs by accident leads only to a *necessary* condition for luck. ACCIDENTAL CHANCE would need to be supplemented with at least a significance condition. That is to say, in the same way as non-significant events occurring by chance are not lucky (e.g., a rockfall in a distant planet), non-significant accidental events are not by luck. If the accidental collision of two rocks affects no one, we would not say that the collision is by luck.

However, for Peter Unger (1968) it is not so obvious that accidentality does not imply significance at all. He has the intuition that "[w]hat we properly regard as an accident, or as accidental, does appear to depend upon our various interests" (Unger 1968: 159). Therefore, there

2 The connection between luck and accidentality can be traced back to Unger (1968), where he offers several diagnoses of Gettier-style cases and defines knowledge in terms of non-accidental belief.
3 This well-known case is by Roderick Chisholm (1977: 105).

are two conflicting intuitions here. On the one hand, there is the intuition that there are events that occur by accident that have no effect on our interests. On the other hand, there is Unger's intuition that we only regard as an accident, or as accidental, events that depend on our interests.

§ E is an accident vs. E is by accident. There is no clash of intuitions. If conflicting intuitions concerning accidentality arise, it is because two different things are being conflated. It is not the same to claim that *E* is an *accident* as to claim that *E* occurred *by accident* (or that it is accidental that *E* occurred). It is obviously true that if *E* is an accident, it is accidental that *E* occurred, but the other direction does not always hold, namely that if it is accidental that *E* occurred, *E* is an accident. Significance, here, makes the difference.

It might be accidental that two objects collide, but we would be reluctant to apply the term 'accident' to a collision if it does not have a positive or negative effect on our interests. For example, the collision of two cars is a paradigmatic instance of accident not only because it is accidental, but also because the fact that two cars collide is something that has great impact on our interests: lives are typically at stake. By contrast, we are more reluctant to call an accident the collision of two atoms at the other end of the galaxy (we would simply say that it is accidental that they collide), and if we call it an accident, it is plausibly because the collision is significant, e.g., for scientific theories (i.e., it affects our epistemic interests).[4] To conclude, *accidents* are partially dependent on our interests, something that is not true of *accidental occurrences* (ACCIDENTAL CHANCE is formulated in terms of accidental occurrence).

§ Objection: uninformativeness. The big problem with ACCIDENTAL CHANCE is that it is completely uninformative. It simply does not state the conditions under which an event occurs by accident. Consequently, to equate chance with accidental occurrence is just to replace an undefined notion by another undefined notion. Therefore, unless more is said about under which conditions the occurrence of an event counts as accidental, ACCIDENTAL CHANCE remains a not very illuminating condition. One could try to make it more informative using probabilistic or modal notions. However, there are reasons not to proceed in that way. For it is not a good strategy to interpret the notion of accidentality in terms of probability or modality so as to formulate an adequate chance condition for luck when the chance condition itself can be put directly in probabilistic or in modal terms. Simplicity constrains us to drop the notion of accidental occurrence.

---

4 It is not easy to find examples of non-significant events, because even events that seem to have no significance at all acquire a little significance by the very act of using them as examples.

1.1.1 *Are Lucky Events Accidents?*

Still, the narrower notion of accident could be rescued to define the notion of luck itself. The idea would be simply to think of lucky events as the sort of events that we call accidents:

- Accidental Luck: *E* is lucky for *S* if and only if *E* is an accident that concerns *S*.

As far as I know, nobody has analyzed luck in this way, but it is interesting to evaluate this definition so as to mark the contrast with Accidental Chance. In principle, we are allowed to define the notion luck in terms of the notion of accident because, as we have seen, an event is an accident only if it has an impact on our interests, i.e., it is somehow significant for us (recall that an event can accidentally occur without its occurrence having an impact on our interests; in that case, the event would not be an accident). As we will see next, although quite intuitive, Accidental Luck is problematic for two reasons.

§ Objection 1: degrees of luck vs. degrees of accidentality. First, while the notion of luck is intuitively gradual, it is not so evident that the notion of accident is gradual. Finding a briefcase with one million inside is luckier than finding a coin, but both events seem to be accidents in the same sense.

Nevertheless, this intuition might not be shared by everyone and namely by those who think that it is possible to provide an account of degrees of accidentality. Such an account could be given, for instance, by taking into account the probability or modality of accidental events. In this way, finding a briefcase with one million inside would be more accidental than finding a coin, because, intuitively, the former is less probable or modally robust than the latter.[5]

However, even if that account were successful, it could not provide a complete explanation of the gradualness of luck. In other words, even if it were true that the notion of accident allows for degrees, its quality of being gradual is not the same quality of being gradual that luck involves. The reason is that the degree of luckiness of an event is not only determined by the degree of probability or the modal profile of the event, but also, intuitively (and desirably), by the degree of significance that the event has for us: finding a briefcase with one million inside is luckier than finding a coin because the former is more significant (positively or negatively) than the latter. Yet, the degree of accidentality of an event does not seem to be a function of its significance.

Although accidents depend on our various interests (and thus to regard an event as an accident it suffices that its accidental occurrence

5 E_1 is less modally robust than E_2 if and only if E_1 would occur in less close possible worlds than E_2.

is somehow related to our interests), the degree of significance of an event does not increase or decrease the degree of accidentality of its occurrence: killing a mouse in a fortuitous car crash is as accidental as killing a person in fortuitous car crash. By contrast, killing a person in a car crash involves a greater degree of bad luck than killing a mouse (because, in general, we ascribe more significance to the life of a person than to the life of a mouse). Thus, if degrees of accidentality make sense, they are not a function of the sort of significance relative to which the notion of luck is sensitive and, therefore, they cannot appropriately account for degrees of luck.

§ Objection 2: not all lucky events are accidents. Second, as Duncan Pritchard (2005) argues, another (more obvious) reason why ACCIDENTAL LUCK is incorrect is, simply, that there are paradigmatic cases of luck that do not involve accidents, e.g., winning a lottery. The reason he gives is that "[i]f one deliberately bought the ticket in question and, say, one self-consciously chose the winning numbers, then it would be odd to refer to the resulting outcome as being accidental" (Pritchard 2005: 126).

Is *intentionality* what makes the difference between the notions of luck and accident? In what follows, I will show that the fact that an action is intentional is necessary but not sufficient to exclude the accidentality of an event that results from that action. The argument aims to shed some more light on the nature of luck. But let us recap first. It should be clear at this point that the notion of luck cannot be reduced to the notion of accident. Two reasons have been given: first, the gradualness of luck is not of the same kind as the gradualness of accidents (if there is any at all); second, there are lucky events that are not accidents.

§ When is the result of an action not accidental? Let us continue. Someone who holds that the intentionality of an action is sufficient to exclude the accidentality of an event would endorse the following conditional:

- If S φ-*es* with the intention of bringing about result R, then if R occurs, R is not an accident.[6]

Is it necessarily true? I do not think so. Imagine that a thief being chased by a police car prays and asks God to cause a rockfall to block the road and to make possible her escape. Imagine that a rockfall suddenly happens. Assuming that God does not exist,[7] is the rockfall an accident? It seems so. Therefore, the conditional is not necessarily true: the thief, by praying to God, has the intention that a

6 Although something like this principle underlies Pritchard's quote above, I am not sure whether Pritchard actually endorses it.
7 If one does not want to grant this assumption, one only has to suppose that the thief believes that, by making the kind of verbal ritual she makes, she will obtain a mental super power to provoke rockfalls.

rockfall happens; the rockfall happens; yet, it is by accident. Conclusion: the intentionality of an action is not sufficient to make an event non-accidental.

Under what conditions can an agent's action prevent an event from being accidental? The answer is to be found, not only in the intentionality of the action, but also in the *causal relevance* it has for the occurrence of the event in question. Compare the case of the thief, where she has the intention that a rockfall happens but her actions (praying to God) are causally irrelevant to the occurrence of the rockfall, with the lottery case described by Pritchard above in which one wins the lottery after having bought a ticket that one has self-consciously chosen. In the former case, the event in question is accidental; in the latter, it is not. Why? Because the causal relevance of the action (buying a lottery ticket) makes a difference on whether the event (winning the lottery) is accidental or not.

Note, however, that it makes a difference only if one, by performing that action, aims to bring about the event (i.e., only if one has the aim of winning the lottery with that ticket). To see this more clearly, imagine that a pilot, who is dancing in the flight deck, unintentionally presses the depressurization button, which causes the crash of the airplane. We would consider the airplane crash an accident even though the pilot's action is causally relevant for its occurrence. This indicates that intentionality and causal relevance are both needed to exclude accidentality. Accordingly, the following conditional seems necessarily true:

- If an action that S has performed with the intention of bringing about result R is causally relevant for the occurrence of R, then R is not an accident.

As I said before, the aim of the argument is to shed some light on the notion of luck. Note, first, that all cases considered are cases of luck. In the lottery case, it is by luck that one wins the lottery. In the thief case, it is by luck that the rockfall happens. In the pilot case, it is by (bad) luck that the airplane crashes. In addition, only the rockfall and the airplane crash are accidents.

This leads us to state the fundamental difference between the notions of luck and accident. Both notions have to do with our interests and both might be explained in probabilistic or in modal terms. The fundamental difference between them is that while an event does not qualify as an accident if an agent performs an action with the intention of bringing it about and the action is causally relevant for the occurrence of the event, it is possible that an event is lucky when an agent acts with the intention of making the event happen and her action is causally relevant for its occurrence (e.g., winning a lottery).[8]

8 To strengthen the intuition that causally relevant intentional action is compatible with a lucky outcome, think about a lottery in which one does not buy a ticket but picks a ball from the lottery drum with a blindfold on.

In this sense, the notion of luck is broader than the notion of accident. Since it seems impossible to conceive a case in which an event is an accident but it is not lucky, we can plausibly conclude that all accidents are lucky, but not all lucky events are accidents.

Let us recap. We started by analyzing a possible way of formulating the chance condition for luck: in terms of the notion of accidental occurrence. But we have seen that such a formulation is unsatisfactory because it is uninformative. We have then considered another possibility: to conceive lucky events simply as accidents, which are events that accidentally occur and that depend on our interests. But we have seen that this equivalence is problematic for two reasons. First, while the degree of luck of an event is a function of the degree of significance that the event has for us, the degree of accidentality of an event (if any) is not. Second, and more importantly, there are lucky events that are not accidents: events that result from actions that are performed with the intention of bringing them about.

1.2 COINCIDENCES

I will briefly contrast luck with another notion that is very similar to the notion of accident: the notion of *coincidence*. According to David Owens (1992), coincidences are inexplicable events in the sense that their constituents are produced by independent causal factors in such a way that we cannot explain why those constituents come together (i.e., there is no close common nomological antecedent of the components of the coincidence or a close nomological connection between them to which we can appeal to explain why the coincidence in question occurs).[9][10]

Consider again the thief case. A rockfall happens in the precise moment when the thief is asking God to make a rockfall happen. A causal process leads to the rockfall. Another psychological causal process leads to the praying. However, there is neither a close common nomological antecedent nor a close nomological connection between them. Thus, the rockfall is a coincidence.

The rockfall is also by luck. Coincidences (which are pairs of events) seem to involve luck. This is obvious and unproblematic. However, may a pair of events involve luck without being thereby a coincidence? Suppose that *S* tosses a coin and lands heads, the result that *S* wished to obtain. The pair of events *S's wishing the coin landing heads* and *S's obtaining the result she wanted* involves luck. Is such a pair of

9 I say 'close' because there can be coincidences with far nomological antecedents or connections (e.g., the Big Bang does not prevent from being a coincidence that you have wished that your favorite team wins the final and that as a matter of fact your team has won the final).

10 Rescher (1995: 215) makes a similar point about fortuitousness. He says: "A conjecture is fortuitous if it involves the concurrent realization of events that are produced by chains of causality operating independently of one another".

events a coincidence? Not so clearly. *S*'s tossing the coin intentionally seems to prevent it from being a coincidence. That is, in the same way as causally relevant intentional action prevents an event from being an accident, causally relevant intentional action seems to prevent a pair of events from being a coincidence. In conclusion, the notion of luck does not seem reducible to the notion of coincidence.

1.3 INDETERMINACY

In a *causally deterministic* world an event *E* that happens at *t* is necessitated as a matter of natural law by antecedent conditions. An idea is that lucky events are events whose occurrence was not predetermined prior to their occurrence. Accordingly, one way of conceiving the chance condition for luck is in terms of *metaphysical indeterminacy*, i.e., indeterminacy that arises from the world rather than from language (semantic indeterminacy) or from one's epistemic position (epistemic indeterminacy):

- Objective Indeterministic Chance: *E* is lucky for *S* only if its occurrence at *t* is not necessitated as a matter of natural law by antecedent conditions.

§ *Objection: luck without indeterminacy*. Even under the assumption of determinism, many would still ascribe luck to people.[11] For example, many would still claim that *S* is lucky to have won the lottery even though we know that determinism holds in such a way that, given the physical state of the balls at t_1 and the laws of nature that govern the behavior of the balls in the lottery drum, it is determined that the ball that makes *S* win the lottery will come out at t_2. Plausibly, attributions of luck are true in spite of determinism (although there will be for sure philosophers who will reject that).

§ *Luck attribution and indeterminacy.* Although Objective Indeterministic Chance does not state a necessary condition for luck, it is still worth analyzing the relation between indeterminacy and luck attribution, as it can provide useful insights into the nature of luck and a curious chance condition.

The first thing to note is that when we participate in a lottery we typically consider the lottery process an indeterministic device even though it is deterministic. Accordingly, a *prima facie* plausible (partial) explanation of luck attribution could be the following: the fact that we attribute luck to a lottery winner is due to the fact that we consider that the process that leads to her winning is indeterministic (even if, strictly speaking, it is deterministic). If this explanation is correct, we should *not* expect people to judge that an event is by luck if the event

11 See Coffman (2007: 389) and Pritchard (2005: 126-127) for an articulation of this kind of objection.

results from a process that they consider deterministic. As we will see next, people might make this kind of deterministic judgments when the relevant events are the outcomes of indeterministic processes (viz., events that are by luck if significant).

To shed light on these issues concerning luck attribution, it is useful to compare two cases by Daniel Dennett (1984: 120), which he introduces with other purposes in the context of a discussion about free will (he aims to show that there are real opportunities). Dennett asks us to consider two lotteries. In lottery *A* players buy their tickets first and then they place them in a quantum-mechanically random system that extracts one ticket, the winner one. In lottery *B* the same system extracts the winner ticket before the tickets are sold.

What are the folk intuitions about the cases? Most people would surely consider that the lottery process of lottery *A* is indeterministic. However, when it comes to lottery *B*, there would surely be many people who would think that the lottery process is deterministic even though it is indeterministic (and, plausibly, even if they were told that it is indeterministic they would still treat it as deterministic). The cause of why so many people would have a deterministic intuition about lottery *B* is that they would think that the winner ticket is already the winner before the tickets are sold.

Let us return now to luck. Imagine that *S* buys tickets of the two lotteries and wins both. As regards lottery *A*, most people would say that *S* is lucky to have won. Concerning lottery *B*, however, people with the deterministic intuition would surely deny that *S* is lucky to have won the lottery, and they would think so despite the fact that the chance of winning is the same in both lotteries. In this way, if luck attributions are sensitive to how people think the world is rather than how it really is, and namely to whether they treat the relevant processes as deterministic or indeterministic, the following condition might be on the right track:

- Subjective Indeterministic Chance: The occurrence of *E* at *t* is lucky for *S* only if the last doxastic state of *S* about *E*'s occurrence before *t* is not that that *E*'s occurrence at *t* is determined to occur.

However, Subjective Indeterministic Chance leads to odd results. If in order for an event to be lucky it is necessary that one does not consider the occurrence of the event as already determined to occur, there will be people who will never have good or bad luck.

Some of the beliefs of very religious people are good counterexamples to Subjective Indeterministic Chance. Consider a person who has a deep belief in God's plan and who consequently believes that everything that happens in her life follows a predetermined course of events. Suppose that this person wins a lottery. She would consider her winning the lottery part of God's plan and, as a result, the condition would imply that this person is not lucky to have won the lottery. However, that is completely counterintuitive: that person is objectively lucky to have won. Thus, the relevant chance condition for luck cannot be formulated in terms of whether the agent considers that the relevant event is determined or not to occur. Luck, so far as chance is concerned, is not a matter of perspective (if by 'perspective' is meant to think the relevant processes that lead to the events in question as being deterministic or indeterministic).

Let us recap. Indeterministic processes are for sure an excellent way of generating lucky events. For example, winning a quantum-mechanically random lottery is, if significant, always by luck. However, we have seen that from the fact that indeterministic processes produce lucky events (if significant) it does not follow that, whenever the process of which an event results is deterministic, the event in question cannot be by luck. Furthermore, although everything that happens in the present were already determined to occur as a matter of natural law by how the world was in the past, many would still regard many events as lucky (e.g., lottery wins).[12]

Besides, as we have seen, it is also an error to think that the lucky status of an event depends on whether we *think* that the process that produces the event is deterministic or indeterministic. Nobody would deny that a person who believes in destiny is lucky to win a lottery. There always seem to be objective facts of the matter about whether an event is lucky.[13] However, before rejecting the view that luck is a perspectival matter, we must evaluate first another chance condition that entails this view.

1.4 SUBJECTIVE PROBABILITY

Nicolas Rescher (1995), in one of the first studies entirely dedicated to the notion of luck, maintains that "[l]uck is the antithesis of reasonable expectation" (Rescher 1995: 35). This claim can be read as the claim that *E* is lucky for *S* only if the occurrence of *E* does not seem to *S* a reasonable thing to expect. This kind of view can be described as perspectival, in the sense that the lucky status of an event for an agent depends on the agent's beliefs (or lack of beliefs) about the event.

More specifically, the extent to which an agent finds reasonable to believe a proposition about the future occurrence of an event (call this belief an expectation of occurrence) can be interpreted in terms of how much *credence* gives the agent to the proposition. The greater credence the agent gives to the proposition that *E* will occur, the more reasonable the agent will find that *E* will occur (i.e., the more she will expect its occurrence).

12 As noted before, not all philosophers will be willing to accept this point.

13 For similar reasons, the following two chance conditions are incorrect: 1) *E* is lucky for *S* at *t* only if *S* considers *E*'s occurrence accidental; 2) *E* is lucky for *S* at *t* only if *S* considered before *t* that *E* was at risk of not occurring.

Degrees of beliefs or credences indicate how confident we are about propositions or how strongly we believe them. A common assumption is that the credences of rational agents can be represented numerically (with real numbers) and that they obey probability calculus. In particular, credences are commonly conceived as *subjective probabilities*. In this sense, the claim that E is lucky for S only if the occurrence of E does not seem to S a reasonable thing to expect can be understood as a chance condition for luck formulated in terms of subjective probabilities.[14] Asbjørn Steglich-Petersen (2010) proposes (but does not endorse) a formulation of a chance condition for luck in these terms:

- Subjective Probabilistic Chance: S is lucky with respect to E at t only if S had a low degree of belief just before t that E would occur at t. (Steglich-Petersen 2010: 366)

§ *Can the lucky status of an event change with our expectations?* Before analyzing the problems of Subjective Probabilistic Chance, it is important to emphasize an implication that the perspectival approach to luck has. This approach leaves room for the possibility that, as far as chance is concerned, one and the same event is lucky for A but not for B. More specifically:

- Perspective: If for A but not for B it is subjectively probable that E will occur, then E is lucky for A but not for B.

Perspective might be what underlies Rescher's intuitions about the following case:

> Big Check
>
> Your secret benefactor's sending you that big check represents a stroke of good luck for *you* even if it is something that *he* has been planning for years. (Rescher 1995: 35; emphasis in the original)

It seems that Rescher, by emphasizing 'you' and 'he', wants to indicate that the big check is a lucky event for you (since your degree of belief that the event would obtain was low) but not for your benefactor (since his degree of belief that the event would obtain was high).

Andrew Latus (2003) explicitly endorses Perspective. Latus formulates a case in which E is bound to occur, but because of being unexpected by S, E is lucky for S but unlucky from the perspective of someone who knew that E was bound to occur:

> Rich Uncle

14 In what follows, I will talk indistinctively about degrees of belief, credences, subjective probabilities or expectations of occurrence (I refer to all these when I talk about perspectives or points of view).

> A nephew receives a big sum of money from his rich uncle despite the fact that in his whole life the uncle has always told the nephew that he would not receive a penny from him. Unbeknownst to the nephew, the uncle, by saying such things, was just testing the nephew's behavior, a test that the nephew has passed, because he has always been a good boy. (Adapted from Latus 2003: 468)[15]

Consider Latus's diagnosis of the case:

> As I have described the case, it would have been unreasonable for [the nephew] to think [he was] likely to inherit the money. But surely there is another perspective from which what happened is not a matter of chance. To someone who knew of [the] uncle's plans and had some idea of [the nephew's] character (and therefore that [he] would almost certainly pass the test), this would not seem a matter of chance at all but something quite predictable. Which, if either, of these perspectives is the correct one when it comes to determining the luck involved in this occurrence? The answer is that there is no need to pick one of these perspectives as the correct one. Luck is, after all, a matter of perspective in at least one sense. It is quite possible for an event to be good luck for one person and bad luck for another. Your good luck in having the single lottery ticket you bought turn out to be the winning one may be another person's bad luck in that the only ticket he didn't buy turned out to be the winning one. What we have discovered by considering the 'rich uncle' example is that luck is also a matter of perspective so far as chance is concerned. (Latus 2003: 468-469)

Cases like BIG CHECK and RICH UNCLE seem to speak in favor of PERSPECTIVE. However, PERSPECTIVE is modeled on a chance condition for luck that says that an event is lucky for an agent at a certain time only if the event was subjectively improbable before that time for the agent (SUBJECTIVE PROBABILISTIC CHANCE). That is, the tenability of the former depends on the tenability of the latter. Is SUBJECTIVE PROBABILISTIC CHANCE tenable?

I do not deny that it has some initial plausibility. For example, it gives a reasonable explanation of why a lottery winner is lucky: the view would appeal to the fact that, before winning, the lottery winner was probably not very confident that she would win. Furthermore, it follows from PERSPECTIVE that winning a lottery represents a stroke

15 As we will see, some commentators consider that cases of this sort cannot be properly called cases of luck but of *fortune*. Besides, note that RICH UNCLE is structurally equivalent to BIG CHECK, which Rescher considers a case of luck. More will be said on this clash of intuitions at the end of this chapter.

of luck not only from the winner's point of view, but also from the point of view of the rest of lottery players: it was subjectively improbable for them that that person would win. Some might find these explanations plausible. However, despite their *prima facie* plausibility, Subjective Probabilistic Chance and hence Perspective face two serious problems.

§ *Objection 1: irrational expectations*. The first problem is that our expectations are not always based on the evidence we have. This fact may lead the defender of Subjective Probabilistic Chance to attribute luck when intuitively there is not. By way of illustration, the view would not ascribe luck to agents who are cognitively biased in certain situations of luck. This is the case of illusions of control, which increase, for instance, the expectations of winning a lottery.[16] Many other cognitive biases like the well-known gambler's fallacy are also good examples of non-epistemic irrational factors that increase expectations of occurrence.[17]

In addition, we can imagine even more extreme cases in which someone with some mental or personality disorder expects in an absolutely irrational way that E, a very significant event, will occur. Suppose that E is a very improbable event whose occurrence no one expects. Suppose, however, that for some extraordinary reason E occurs. Under the irrational perspective of that person, E was bound to occur, so that E would not be lucky for her. That is, it would follow from Subjective Probabilistic Chance that an event that happens by sheer coincidence is not lucky as a matter of irrational expectation.[18]

§ *Objection 2: no thought, no luck*. Another problem for Subjective Probabilistic Chance is that it blocks the possibility of an event being lucky for an agent when the agent does not entertain any thought about the event. Consider the following case.

> Bloody Dictator
>
> A fascist dictator decides to kill all people with left-wing views, but he considers 'ethical' to save two of them. To that aim, he puts pieces of paper with the names of all

16 In a series of experiments, Ellen J. Langer (1975) observed that participants were more prone to expect winning a lottery when 1) they were given the option of choosing the lottery ticket and 2) when the ticket was familiar to them. Familiar objects and choices are factors typically involved in situations of control. For this reason, although lotteries are chance situations, the participants were more prone to expect winning when they were familiar with their tickets or when they had the possibility of choosing them: in both cases they had the impression that they could control the outcome of the lottery.

17 The gambler's fallacy arises when a subject comes to believe that certain outcome of a game of chance will be the case after seeing that a different outcome has repeatedly been the case. In general, the fallacy refers to the belief that the probability of a randomly produced event is influenced by independent events.

18 The point about irrationality is also made by Steglich-Petersen (2010: 366).

> the suspects of having left-wing views written on them in a lottery drum. His idea is to extract at random two names, the ones who will not be murdered. The first piece of paper that the dictator extracts corresponds to Charles, a communist spy who is aware of the plans of the dictator. The second piece of paper corresponds to Michael, an anarchist who has awoken from a one-year coma just a couple of minutes before the dictator extracts his name. The dictator came to power six months ago, so Michael does not know anything about the actual political situation. He still thinks that he lives in a democracy.

Are Charles and Michael lucky to have survived? According to SUBJECTIVE PROBABILISTIC CHANCE, Charles is lucky to have survived, since he was aware of the macabre lottery and consequently not very confident that he would survive. By contrast, Michael does not know anything about the dictator and his macabre lottery. Moreover, after the terrible car accident that Michael suffered one year ago his coming out of the coma makes him very confident that he will survive *before* the macabre lottery draw takes place. Accordingly, SUBJECTIVE PROBABILISTIC CHANCE implies that Michael's survival is lucky from the dictator's perspective but not from his point of view. This result is anything but intuitive. In addition, the lucky status of his survival would change once Michael gets access to the relevant information. The intuition, however, is that Michael has always been lucky independently of what he thinks about it.

In conclusion, the view that luck, as far as chance is concerned, is a perspectival matter leads to highly counterintuitive results, to the extent that it makes its core chance condition (SUBJECTIVE PROBABILISTIC CHANCE) unnecessary for luck. In the next section, I will analyze several formulations of the chance condition for luck in terms of epistemic probabilities and Steglich-Petersen's chance condition in terms of the notion of being in a position to know.

1.5 EPISTEMIC PROBABILITY

Another way to formulate the chance condition for luck is in terms of epistemic probabilities, which measure the degree of support of one's evidence to hypotheses about future occurrence of events.[19] Steglich-Petersen (2010), who has recently proposed the view that

19 As Mellor (2005: 81) explains, epistemic probabilities are conditional probabilities (the probability of *A* given *B*) as long as they express probabilistic relations between two propositions: the probability of certain proposition (a hypothesis) given another proposition (certain piece of evidence). Furthermore, as he explains, epistemic probabilities are objective because whether certain piece of evidence confirms or disconfirms a hypothesis is an objective matter and not a merely matter of opinion (Mellor 2005: 8). For example, it is not a matter of opinion that the data collected by the LHC at CERN make extremely probable the existence of the Higgs boson.

luck is an epistemic notion, explores several alternative chance conditions in terms of epistemic probabilities before formulating his own one. To shed more light on the nature of luck, let us analyze them and let us assess why they fail. The first condition that he considers introduces an objective relation of evidential support:

- Epistemic Probabilistic Chance 1: *E* is lucky for *S* at *t* only if, given the evidence available to *S* just before *t*, there was a large chance that *E* would not occur at *t*.

I noted before that Subjective Probabilistic Chance is problematic because if an agent entertains no thought at all or has no expectation about the occurrence of an event, the event does not count as lucky for the agent even if it represents a blatant stroke of luck for her. Steglich-Petersen (2010: 367) argues that Epistemic Probabilistic Chance 1 has the same problem: *E* might be lucky for *S* even if *S* has no prior evidence about the occurrence of *E*. For this reason, he proposes these other two chance conditions:

- Epistemic Probabilistic Chance 2: *E* is lucky for *S* at *t* only if, for all *S* knew just before *t*, there was a large chance that *E* would not occur at *t*.

- Epistemic Modal Chance: *E* is lucky for *S* at *t* only if, for all *S* knew just before *t*, there was a wide class of close possible worlds in which *E* would not occur at *t*.

Steglich-Petersen thinks that these two similar conditions are problematic because they offer a wrong diagnosis of paradigmatic Gettier-style cases. Suppose that *S* has strong evidence for the proposition that *E* will occur and forms the belief that *E* will occur. *E*, however, is an event that will probably not occur but *S* expects its occurrence because the evidence is misleading. Suppose that by an extraordinary coincidence *E* ends up obtaining so that *S*'s belief becomes true. It is by luck that *S*'s belief is true but for all she knew before the occurrence of *E*, *E* would have occurred. Therefore, Epistemic Probabilistic Chance 2 and Epistemic Modal Chance do not deliver the correct diagnosis that the belief is luckily true, or so argues Steglich-Petersen.

The objection can be extended to all cases of luck (it does not only apply to cases of veritic epistemic luck).[20] Recall Bloody Dictator. For all that Charles knew before the macabre lottery draw, he would not survive (he knew that his name was in the lottery drum). On the contrary, for all that Michael knew, he would survive (he did not even know that a dictator had recently come to power). Since both Charles and Michael win the lottery, it is intuitively correct to say that it is by luck that both are still alive. The problem is that it follows from

20 Veritic epistemic luck will be defined in chapter 4.

EPISTEMIC PROBABILISTIC CHANCE 2 and EPISTEMIC MODAL CHANCE that Michael is not lucky to have survived. Given what Michael knew before winning the lottery, there was a large chance that he would survive (alternatively, there was a wide class of close possible worlds in which he would survive). After all, since he knew absolutely nothing about the dictatorship, he had no reason to think that he would die.

1.5.1 *Is Luck an Epistemic Notion?*

In order to remedy the defects of EPISTEMIC PROBABILISTIC CHANCE 2 and EPISTEMIC MODAL CHANCE, Steglich-Petersen thinks it plausible to resort to the notion of *being in a position to know*. In particular, he formulates the following chance condition for luck:

- WEAK EPISTEMIC CHANCE: *E* is lucky for *S* at *t* only if, just before *t*, *S* was not in a position to know that *E* would occur at *t*.

According to Steglich-Petersen, being in a position to know that *E* will occur at *t* "simply means that the agent is in such an epistemic position with regard to the occurrence of *E*, that *if* she takes up the belief that *E* will occur, she will know that *E* will occur"(Steglich-Petersen 2010: 368). In Gettier-style cases, agents are not in a position to know the relevant propositions. In BLOODY DICTATOR, not even Charles is in a position to know that he will survive: knowing the dictator's intentions and that his name is in the lottery drum does not put him in a position to know that he will survive. At most, he knows that his chances of surviving are very low. The proposal has, therefore, some plausibility.

§ *Being in a position to know*. Steglich-Petersen gives a very rough characterization of what it takes to be in a position to know a proposition. Since the notion plays an essential role in WEAK EPISTEMIC CHANCE, more should be said on what conditions must an agent satisfy for being in a position to know some proposition. Sven Rosenkranz (2007: 69), following Timothy Williamson (2000: 95), states three necessary conditions for being in a position to know a proposition *p*:

(1) *p* is true.

(2) *S* is physically and psychologically capable of knowing *p*.

(3) Nothing stands in *S*'s way of successfully exercising these capabilities.

What condition(s) do agents fail to satisfy in cases of luck? In cases of luck (1) is always satisfied, as the relevant proposition <*E* will occur at *t*> is always true in them. Accordingly, we can specify WEAK

Epistemic Chance as the condition that *E* is lucky for *S* at *t* only if, just before *t*, ¬(2) *S* was not physically and psychologically capable of knowing that *E* would occur at *t* or ¬(3) something stood in *S*'s way of successfully exercising these capabilities.

In addition, Rosenkranz explains that there are two ways of reading the expression 'capable of knowing' in (2): a wide and a narrow sense. According to Rosenkranz, one is capable of knowing *p* in the wide sense if one meets all the physical and psychological preconditions for acquiring a decision procedure for *p*. In the narrow sense, one must already possess such a decision procedure:

> (2)* *S* possesses a decision procedure for *p.*

Steglich-Petersen says: "[w]hen an agent is in a position to know that p, all the ingredients for knowledge are in place, apart from belief" (Steglich-Petersen 2010: 368). Having a decision procedure for the proposition <*E* will occur at *t*> is one of the ingredients for knowing it. Therefore, if an agent fails to possess such a decision procedure, she is not a position to know the proposition and, according to Weak Epistemic Chance, *E* is lucky for the agent when *E* occurs. Additionally, Rosenkranz argues that, once we assume (2)*, (3) can be specified in the following way:

> (3)* The enabling conditions for *S*'s successful implementation of the relevant decision procedure are *de facto* met.

(1), (2)* and (3)* are necessary for being in a position to know, but there does not seem to be any reason not to consider them jointly sufficient as well. I will adopt then this clear conception of the notion of being in a position to know to shed light on Weak Epistemic Chance. Accordingly, the condition can be reformulated as follows:

- Weak Epistemic Chance: *E* is lucky for *S* at *t* only if, just before *t*, ¬(2)* *S* did not possess a decision procedure for the proposition that *E* would occur at *t* or ¬(3)* the enabling conditions for *S*'s successful implementation of the relevant decision procedure were not *de facto* met.

In cases of luck agents allegedly fail to satisfy (2)* or (3)* (or both). Contra this view, I will argue that both (2)* and (3)* may be satisfied by an agent and the relevant event may still count as lucky for the agent, i.e., I will show that the condition stated by Weak Epistemic Chance is not necessary for luck.

§ *A counterexample to the necessity of* Weak Epistemic Chance. The following case shows that one might be in a position to know that a lucky event will occur:

> Fair Lottery Player

> John is having serious financial problems. Desperate as he is, he thinks it would be a good idea to buy a Megalotto ticket, which on that occasion has a 50-million jackpot. The peculiarity of Megalotto is that a random system extracts the winner ticket before the tickets are sold. After that, lottery workers put the tickets into envelopes that are subsequently randomly distributed.
>
> John goes to buy a lottery envelope to the grocery store of his friend Jim. Jim is aware of John's financial problems, so he tells John: "Look, buy this envelope: the winner ticket is inside". John asks Jim how can he know that, to which Jim answers that he knows it because his cousin, a lottery worker, has told him that he has put the winner ticket inside that specific envelope.
>
> However, Jim's cousin was just playing a joke on Jim. The strict lottery system prevents workers from knowing where the winner ticket is. Suddenly, Jim receives an emergency call and must run to the hospital immediately. He asks John to stay in charge of the store while he is absent. And there he is John, before the envelope, having the chance of checking whether the winner ticket is inside. After a few seconds of hesitation, John deliberately refuses to open the envelope. "Honesty above all", he thinks. When Jim returns, John buys the envelope. Luckily, the winner ticket is inside and John wins the lottery.

John is in a position to know that he will win the lottery. The decision procedure is as simple as opening the envelope, accessing conclusive evidence and leaving money on the counter with a note saying "Jim: here's the money for the ticket (and a little extra)" (i.e., condition (2)* holds). However, John deliberately refuses to act in that way. In addition, the enabling conditions for that procedure are *de facto* met: namely, Jim is absent, John has hands, a properly working visual system, sufficient background knowledge of Megalotto to identify the winner ticket, and so on and so forth (i.e., condition (3)* holds). In addition, John will win the lottery (i.e., condition (1) is satisfied). Therefore, John is in a position to know that he will win. Accordingly, Weak Epistemic Chance judges that this is not a case of luck. However, it is by luck that John wins, as it is by sheer coincidence that the winner ticket is inside the particular envelope in which Jim's cousin said the winner ticket was. This means that the condition stated by Weak Epistemic Chance is not necessary for luck.

§ Possible reply. In reply, one could argue that the fact that John is honest and, in particular, the thought "Honesty above all" has as a consequence that (3)* is not satisfied (the enabling conditions for John's successful implementation of his decision procedure would

not be met). If this is correct, John is not in a position to know that the winner ticket is inside the envelope, consistently with what Weak Epistemic Chance says.

§ *Rejoinder*. Although it is a little strained to say that John is not in a position to know that the winner ticket is inside the envelope because he is honest, we must take the objection seriously. As a response, the case can be reformulated by replacing John's honesty with whatever reason *R* does not entail the negation of (3)*. Moreover, the case can be reformulated in such a way that John decides not to open the envelope for no reason, e.g., in such a way that everything required for knowing the target proposition is in place but John decides on a whim not to open the envelope.

In any case, leaving the details aside, there is a more general point to be made: the fact that the *enabling* conditions for an agent's successful implementation of a decision procedure are met should not prevent the fact that the agent decides freely not to implement the decision procedure (otherwise, I do not see in which sense the relevant conditions can be properly said to be 'enabling'). In other words, (3)* should be interpreted in such a way that it is compatible with free choice. In view of these considerations, we cannot but conclude that one might be in a position to know that a lucky event will occur without thereby knowing that it will occur. Luck, in this way, is not an epistemic notion.

Let us recap. Several epistemic versions of the chance condition for luck in terms of epistemic probabilities have been analyzed and rejected. On the other hand, special attention has been paid to Steglich-Petersen's view (2010) according to which the lucky status of an event depends on one's epistemic position. A counterexample has shown that being in a position to know that certain event will occur does not prevent the event from being lucky. The moral we can draw is that, sometimes in our everyday lives, we find ourselves in an excellent position to know the future, but we can deliberately refuse to know it thus exposing us to the whims of luck. As we will see in the next section, lucky events are events that occur but that were at risk of not occurring, where the relevant notion of risk is to be understood in modal terms.

1.6 risk

Luck and risk are closely related. Many of the luckiest events we can imagine occur in situations where there is a big amount of risk involved. Being the only survivor in a plane crash or winning roulette after betting one's life savings on one spin are examples of very lucky events that occur in extremely risky situations. In this section, I will investigate this link. First, I will distinguish two senses of risk.

On the one hand, there is a sense of risk that has to do with the possible occurrence or nonoccurrence of an *event*, as when we say that there is risk that a ball on the tip of a cone will fall. Let us call this sense of risk the *event-focused* sense of risk. On the other hand, in addition to affirming or denying that events are at risk of occurring, we also affirm or deny that *agents* are at risk with respect to events, as when we say that a child is at risk with respect to a possible explosion of the WWII grenade that he has just picked up from the ground and with which he is enthusiastically playing. Let us call this sense of risk the *agent-focused* sense of risk. How can these two senses of risk be defined?

1.6.1 *Agent-focused Risk as Lack of Control*

A definition of the agent-focused sense of risk should give an answer to the following question: what kind of relation must an agent have with an event in order for that event to be safe for the agent? I propose that the relation is a relation of *control*: the occurrence of an event is safe for an agent just in case the event is under the control of the agent. The agent-focused sense of risk can be accordingly defined as follows:

- Agential Risk: *S* is at risk with respect to a significant event *E* if and only if *S* lacks control over *E*.

An account of the notion of control will be given in chapter 3 and two *senses* in which an agent might be at risk with respect to an event will be specified. In chapter 3, it will also be explained why the event-focused and agent-focused senses of risk can come apart. For the moment, note that Agential Risk is a technical notion that slightly departs from the ordinary notion of risk because it allows that significant events that have positive effects on an agent count as risky for the agent. As the notion of risk is ordinarily used, however, risky events are events that bear bad consequences. The difference between Agential Risk and the ordinary notion of risk will be discussed in chapter 2. For the purpose of analyzing the chance condition for luck, the discussion will focus, for the moment, on the event-focused sense of risk. But before entering into details, let me state the hypothesis which I will attempt to demonstrate in this first part of the dissertation:

- *S* is lucky with respect to an event *E* that is significant for *S* just in case *S* is at risk with respect to *E* and *E* occurs but was at risk of not occurring.

That is, my hypothesis is that both senses of risk (agent-focused and event-focused risk) are required to define the notion of luck.

1.6.2 *Event-focused Risk*

The event-focused sense of risk can be defined in two ways: in probabilistic or in modal terms. Let us begin with the probabilistic conception.

1.6.2.1 *High Probability of Occurrence*

Consider the following definition of the event-focused sense of risk:

- Objective Probabilistic Risk: *E* is at risk of occurring at *t* if and only if there is high probability that *E* will occur at *t*.[21]

The kind of probabilities that are relevant in Objective Probabilistic Risk are *physical probabilities* or *chances*. Physical probabilities are the kind of probabilities posited by scientific theories, e.g., the probability that an atom of radium decays after 1601 years or the probability that one remembers four digit numbers correctly (Mellor 2005: 10). The values of physical probabilities are neither determined by scientific evidence, nor by degrees of belief, but by features of the world. This is why most philosophers call them *objective probabilities* (hereafter, I will use that expression to refer to them; alternatively, I will simply call them *probabilities*). In this sense, the risk of developing cancer by having a diet based on meat is greater than the risk of developing cancer by having a diet based on vegetables because the probability of the former is higher than the probability of the latter. Objective Probabilistic Risk explains risk in these terms. Finally, the probabilities that are relevant here are non-trivial probabilities (i.e., probabilities other than 1 or 0). The reason is that if an event has probability 1 of occurring we would not say that it is at high risk of occurring, we would rather say that it is a certainty that it will occur.

1.6.2.2 *Close Possibility of Occurrence*

In general, in most situations that we would call risky there is always some significant event that could easily have occurred. By contrast, in most situations that we would call safe things could not have been otherwise. Accordingly, the modal interpretation of the event-focused sense of risk can be roughly put as follows:

- *E* is at risk of occurring at *t* if and only if *E* could easily occur at *t*.[22]

In order to shed some light on this modal interpretation of the event-focused sense of risk, let me bring together two ideas by Timothy Williamson (2009) and Boris Kment (2006):

---

21 For risk of nonoccurrence: *E* is at risk of not occurring at *t* if and only if there is low probability that *E* will occur at *t* (or high probability that *E* will not occur at *t*).

22 For risk of non-occurrence: *E* is at risk of not occurring at *t* if and only if *E* could not easily occur at *t*.

- "Safety is a sort of local necessity" (Williamson 2009: 14).

That is, an event is risky (not safe) just in case its occurrence fails to be necessary in some restricted sense of necessity. Here is Kment's idea:

- "How easily a proposition *P* could have been true is a matter of how much the worlds in which *P* is true depart from actuality. The closer the closest *P*-worlds are, the more easily *P* could have been true" (Kment 2006: 254).

§ *Kment's analysis of the notion of necessity*. Kment's analysis of the notion of necessity is especially useful for our purposes. He thinks that necessity is a special way of being true, namely, "necessary truths are those propositions whose truth is secure and inexorable in a way in which the truth of contingent propositions is not" (Kment 2006: 253). The difference between necessarily true propositions and contingently true propositions is thus qualitative, but Kment also thinks that necessity and possibility come in degrees: the more necessary a true proposition, the less possible its negation. These two ideas lead Kment to the view that there is a *scale of inexorability* where different truths occupy different positions as a function of their degrees of possibility or of necessity.

When is a proposition necessary then? According to Kment, a proposition is necessary just in case its degree of necessity has a value above a specific point in the inexorability scale, a point that he calls the *indifference point*. How can we determine whether the degree of necessity of certain proposition has reached that point? We can determine it, for instance, if we can apply the expression 'could not have been otherwise' (or analogous expressions) to the proposition in certain context.

Is there really such an indifference point? Kment thinks that a proof of its existence is the fact that once we can say in certain context that two different propositions could not possibly have been the case, it is pointless to ask which one could more easily have been the case. That is, there is a point beyond which differences in degree of inexorability are irrelevant in the context.

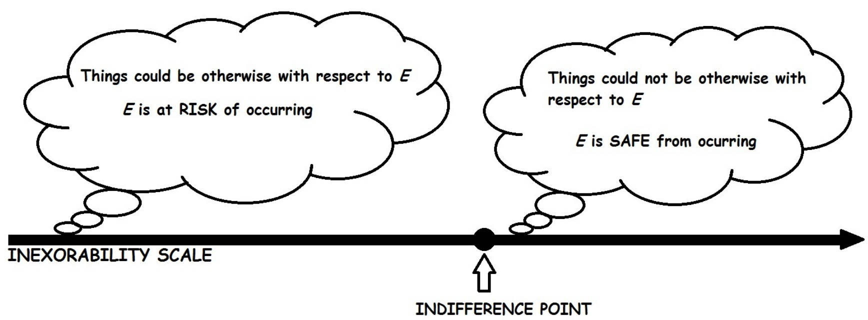
Interestingly, Kment explains that the location of the indifference point in the inexorability scale varies from context to context, i.e., 'could not have been otherwise' expresses different degrees of inexorability in different contexts. He gives two examples: in philosophical contexts, the expression usually refers to metaphysical necessity (which is a very demanding sense of necessity), while in ordinary contexts, which are not so demanding, the expression indicates a lower degree of inexorability, as when we say that Roger Federer's victory against the worst tennis player in the world could not have been otherwise.[23]

23 Kment aims to account for the differences between nomological, metaphysical and conceptual necessity. His view is that each sort of necessity has a different indiffer-

§ *Kment and Williamson's ideas combined*. We can elaborate Williamson's idea that safety is a sort of local necessity using Kment's framework. Imagine that you are in the middle of the Sahara in an extremely hot day. Neither there is risk that you are bitten by a shark, nor that you are struck by a lightning. In this context, both events could not possibly be the case (they are safe from occurring). That is, there is no risk that these eventualities occur because it is *necessary* that they do not occur relatively to your specific situation.

The extent to which it is necessary that they do not occur is of course not the extent to which a proposition is necessary in philosophical contexts. The context described is an ordinary context and thus much less demanding than philosophical contexts, so the degree of inexorability is much lower than the degrees of inexorability that metaphysical, nomological or conceptual necessity require. Compare these two claims: "John is safe from being hit" (when uttered in a panic room) and "John is safe from not being John" (when uttered in a philosophical context). The sort of necessity that these two claims involve is clearly different. The first claim involves practical necessity, whereas the latter involves metaphysical necessity (it is metaphysically necessary that John is John).

In sum, when an event *E* is safe from occurring, the nonoccurrence of *E* reaches an indifference point in the inexorability scale, a point that is set by the context and above which we can only say, given how things actually stand, that *E* could not possibly occur, i.e., that things cannot be otherwise with respect to *E*. If we cannot make such a claim, if things could be otherwise, then *E* is at risk of occurring. The degree of risk of occurrence of an event is thus the degree of possibility that it has of occurring. The following is a simple representation of the inexorability scale that Kment uses to distinguish between contingency and necessity and that I use to distinguish between risk and safety:



We see now more clearly that the indifference point is the point before which an event is at risk of occurring and beyond which the event is

ence point in the *same* inexorability scale, being conceptual necessity the one with a more distant indifference point.

safe from occurring. Beyond that point, it makes no sense to say in a given context that things could be otherwise with respect to the occurrence of *E*.

Following Kment, it is useful to adopt a possible-worlds framework and to consider that the degree of possibility of an event (how easily it could occur) is a function of the closeness of the possible worlds where it occurs. In addition, it is useful to assume that possible worlds (and close possible worlds among them) can be grouped using proportions (e.g., the proportion of close possible worlds in which *E* would occur).[24] A derived assumption of this way of thinking is that proportions can have different sizes.[25] Accordingly, we can give the following modal definition of the event-focused sense of risk:

- OBJECTIVE MODAL RISK: *E* is at risk of occurring at *t* if and only if *E* would occur at *t* in a large enough proportion of close possible worlds.[26]

Why does OBJECTIVE MODAL RISK include the expression 'in a large enough proportion of close possible worlds' instead of, say, 'in most close possible worlds'? Because the indifference point is set by pragmatic factors and consequently it might shift from situation to situation, which means that there is no fixed proportion of close possible worlds in which an event would have to occur to be considered at risk of occurring. In most situations, we would not say that an event is at (significant) risk of occurring unless the proportion of close possible worlds in which it would occur is a large one. Plausibly, this applies to mundane events such as raining, someone dropping a coffee cup or a bulb blowing. However, we sometimes consider that there is significant risk that an event occurs when the relevant proportion of close possible worlds in which it would occur is not large. For example, suppose that you play Russian roulette with one bullet in the chamber of a revolver with a 6-shot capacity. Presumably, given the low probability of being shot (approximately 16%), you would not die in a large proportion of close possible worlds. Still, we would intuitively say that there is risk that you die. Examples of this sort indicate (contrary to what was stated at the beginning of this section) that an event can be at risk of occurring even when its occurrence is not *easily* possible.[27]

---

24 Sometimes, instead of using the expression 'proportion of close possible worlds' I use the expression 'class of possible worlds' (e.g., the class of close possible worlds in which *E* would occur) without implying with that any difference.

25 Sometimes, instead of saying 'in a large proportion of possible worlds' I simply say 'in most possible worlds'. No underlying difference is implied.

26 For risk of nonoccurrence: *E* is at risk of not occurring at *t* if and only if *E* would not occur at *t* in a large enough proportion of close possible worlds.

27 Peacocke (1999), Sainsbury (1997) and Williamson (2000) analyze risk in terms of easy possibilities, or at least they seem to assume such a conception, which I consider inaccurate. See section 1.6.4 for an account of what fixes the indifference point

§ *Two legitimate conceptions of event-focused risk*. To recap, there are two possible ways of conceiving the event-focused sense of risk: a probabilistic way (Objective Probabilistic Risk) and a modal way (Objective Modal Risk). Which one is the correct way of conceptualizing risk? Both, I would say. Objective Probabilistic Risk fits better the notion of risk that is used in scientific and technical contexts, where the risk that an event has of occurring is usually determined with scientific models that calculate, for example, the objective probability of occurrence of that type of event in a long sequence of instances.

Objective Modal Risk, by contrast, fits better the ordinary notion of risk. In everyday life, when it comes to assessing the risk that an event has of occurring, we resort to our cognitive capacity to handle subjunctive conditionals. The judgments delivered by this capacity are arguably less precise than those delivered by a scientific model. However, what might be regarded as a defect is in fact a virtue, because, on the other side of the coin, this allows us to make lot of true risk ascriptions quickly and on the basis of impoverished evidence, something that has an adaptive value.

Of course, this does not mean that we must avoid doing probabilistic calculations to assess the risk of a given situation. Probabilistic calculations can certainly help us to determine whether there is risk that an event will happen and, more importantly, they can guide us when counterfactual thinking delivers incorrect judgments.[28] After all, our counterfactual cognitive capacities are truth-conducive but no infallible.

In sum, Objective Probabilistic Risk and Objective Modal Risk capture two complementary sides of the event-focused sense of the notion of risk. As we will see next, they serve to formulate two different types of objective chance conditions for luck (in terms of objective probabilities and in modal terms) and, in this sense, they help us to conceive the phenomenon of luck as an instance of the more general phenomenon of risk. Nevertheless, although the notions of risk that underlie these two types of chance conditions are both legitimate, I will show that only chance conditions for luck formulated in terms of modal event-focused risk can explain all cases of luck correctly. In particular, I will show that objective probabilistic chance conditions cannot account for cases of *highly probable lucky* events.

1.6.3 *Objective Probability*

According to Objective Probabilistic Risk, E is at risk of occurring at t if and only if there is high probability that E will occur at t. Intu-

---

in a given context and for an explanation of why our intuitions about risk vary accordingly.

28 For example, probabilistic calculation allows to avoid committing errors such as the gambler's fallacy (see fn. 17).

itively, one's winning a fair lottery was at risk of not occurring in the probabilistic sense because, prior to winning, there was high probability that one would lose. Winning a fair lottery is also a paradigmatic case of luck, so there is good reason to formulate a chance condition for luck in terms of Objective Probabilistic Risk:

- Objective Probabilistic Chance: E is lucky for S at t only if, prior to the occurrence of E at t, there was low objective probability that E would occur at t.

One might prefer to use conditional probabilities to formulate the relevant chance condition:

- Conditional Probabilistic Chance: E is lucky for S at t only if, prior to the occurrence of E at t, there was low objective probability conditional on C that E would occur at t.[29]

C is whatever condition one uses to calculate the probability that E will occur. For example, the unconditional probability that Messi will score at the match is high but given C (the fact that he is injured) the probability that he will score is low. Suppose that Messi ends up scoring by luck. Conditional Probabilistic Chance helps to explain why it is by luck that Messi scores: he was injured and therefore it was not very probable that we would score.

Although objective probabilistic chance conditions explain most cases of luck, I will argue that there might be high (conditional or unconditional) objective probability that E will occur at t and yet E might occur by luck at t. That is, I will show that there are highly probable lucky events and hence that probabilistic chance conditions are not necessary for luck.

1.6.4 *Modality*

According to Objective Modal Risk, E is at risk of occurring at t if and only if E would occur at t in a large enough proportion of close possible worlds. Intuitively, one's winning a fair lottery is at risk of not occurring in the modal sense because one would not win in most close possible worlds. Winning a fair lottery is also a paradigmatic case of luck, so there is good reason to formulate a chance condition for luck in terms of Objective Modal Risk.

In what follows and to that aim, I will discuss three competing proposals of modal chance conditions, which are very similar at first sight but that, nonetheless, present subtle but significant differences. By critically comparing them, we will gain deeper insight into the modal nature of luck and, more importantly, we will be in a position to combine several of these subtle differences in our own reinforced modal chance condition. Here are the three conditions:

29 Baumann (2012) assumes this kind of chance condition for luck.

- MODAL CHANCE 1: *E* is lucky for *S* only if *E* occurs in the actual world but does not occur in a wide class of the nearest possible worlds where the relevant initial conditions for *E* are the same as in the actual world. (Pritchard 2005: 128)
- MODAL CHANCE 2: Where *t* is a temporal interval just before *t**, *S* is lucky with respect to *E* at *t** (...) only if *E* does not occur at *t** in at least half the possible worlds obtainable by making no more than a small change to the actual world at *t*. (Coffman 2007: 390)
- MODAL CHANCE 3: *E* is lucky for *S* only if *E* is chancy. *E* is chancy if it occurs in the actual world at *t**, but it fails to occur in a large enough proportion of possible worlds obtainable by making no more than a small change to the actual world at *t*, where *t* is a temporal interval just prior to *t**. (Levy 2011: 17)

The three conditions account for a very simple idea: lucky events are events that were at serious risk of not occurring in the sense of OBJECTIVE MODAL RISK but that for some reason end up occurring. In what follows, I will address the following three questions:

A. What are 'initial conditions'?

B. Does the modal chance condition for luck need a clause on relevant initial conditions?

C. In what proportion of close possible worlds should an event fail to occur in order for its actual occurrence to be by luck?

A. What are 'initial conditions'? In physics, the expression 'initial conditions' is used to describe the state of a physical system at an initial time from which the system subsequently evolves. In MODAL CHANCE 1, Pritchard uses the expression in a more mundane way, namely to refer to any contingent factor of certain situation that is relevant for the occurrence of an event at a certain time.[30] In view of this use, it would be illusory to expect no vagueness in the specification of the initial conditions.

In some cases, we can identify with precision the initial conditions, especially in cases where the relevant events can be explained in scientific terms (e.g., the initial conditions for the outcome of a coin toss are the position, configuration, momentum, and angular momentum of the coin at the beginning of the free fall motion).[31] In some other

30 One possibility is to understand the kind of relevance at issue as explanatory relevance. In this way, the expression 'initial conditions' in Pritchard's MODAL CHANCE 1 would refer to the set of factors that contribute to the explanation of why the event occurs.

31 See Strzałko *et al.* (2008).

cases, especially in those where there is great agent intervention involved (e.g., the outbreak of a war or the breaking-off of a relationship), the relevant initial conditions are not so easy to specify. It is in these cases where we have no other option but, borrowing Pritchard's words, "[to] let our intuitions guide us as to what it should be read as demanding as regards particular scenarios" (Pritchard 2005: 132).

B. Does the modal chance condition for luck need a clause on relevant initial conditions? If specifying initial conditions is such a complicated and imprecise task, why do we need a clause on initial conditions in our modal chance condition for luck? As Pritchard (2005: 131-133) suggests, one could avoid complications by simply dropping the clause and taking into account the unrestricted class of possible worlds.[32]

This is exactly the path that E. J. Coffman and Neil Levy follow. Rather than demanding that the actual initial conditions for *E* must be the same in close possible worlds (where *E*, in turn, must not occur), Coffman and Levy's chance conditions (respectively MODAL CHANCE 2 and 3) simply state that *E* must not occur in close possible worlds obtained by making no more than a small change to the actual world just before *E* occurs.

Coffman and Levy's move is not unjustified, at least not from an intuitive point of view. After all, as Pritchard (2005: 132) points out, the class of close possible worlds will tend to be dominated by worlds in which the relevant initial conditions as we intuitively understand them are the same as the initial conditions of the actual world.

§ The need of a clause on initial conditions. However, Coffman and Levy's requirement that the relevant close possible worlds are to be obtained by making a small change to the actual world *just before* (or at an interval of time *just prior* to) the target event occurs leads to incorrect diagnoses of some cases of luck. Consider the following case:

> IMPULSIVE PLAYER
>
> In order to increase the emotion of the lottery, the organizers decide to permit last-minute ticket purchases. Tim decides on a whim to buy a lottery ticket just before one of the balls falls through the hole. He instantly chooses a number he likes, pays and, luckily, wins the lottery. Tim thinks: "I still don't understand why I had such a sudden whim to buy a lottery ticket. I hate gambling!".

Intuitively, it is by luck that Tim's ticket turns out to be the winner. The general hypothesis of modal chance conditions for luck is that an event is by luck only if it occurs in the actual world but could easily have not occurred. In IMPULSIVE PLAYER, there are two events that

32 Pritchard keeps the clause because he thinks (correctly) that it makes some explanatory work in clarifying how we understand luck.

could easily have not occurred: Tim's decision to buy that specific lottery ticket and Tim's lottery ticket turning out to be the winner. One might want to explain the fact that it is by luck *that Tim wins the lottery* by appealing to 1) how easy could have been that he decided not to buy a lottery ticket (or that specific ticket) and to 2) how easy could have been that another ticket turned out to be the winner. However, the fact that it is by luck *that Tim's ticket turns out to be the winner* is only explained by the fact that that another ticket could easily have been the winner. To see this, note that when we assess whether it is by luck that Tim's ticket turns out to be the winner we hold fix Tim's decision to buy the ticket he likes.

My contention is that Modal Chance 2 and 3 cannot give the latter explanation. As the case is described, Tim's ticket turning out to be the winner and Tim's purchase of the ticket are *almost* simultaneous. Therefore, close possible worlds obtained by making a small change to the actual world *just before* Tim's ticket turns out to be the winner are worlds in which Tim decides not to buy a lottery ticket (since he hates gambling, he could easily have not bought it). If so, Modal Chance 2 and 3 would be compelled to explain the fact that it is by luck *that Tim's ticket turns out to be the winner* by appealing to the fact that in close possible worlds Tim does not buy a ticket (because he decides not to). This explanation is obviously incorrect.

By contrast, a clause on initial conditions allows to explain the case in a natural way. As noted before, when we assess whether it is by luck that Tim's ticket is the winner we intuitively hold fix Tim's decision to buy the ticket he likes. In this way, the fact that Tim decides to buy that ticket is intuitively part of the relevant initial conditions for his ticket turning out to be the winner. Modal Chance 1 says that luck must be evaluated only with respect to what happens in close possible worlds in which the initial conditions are the same as in the actual world. In most of those possible worlds, Tim buys that ticket and another ticket turns to be the winner, which explains why it is by luck that Tim's ticket turns out to be the winner in the actual world. A clause on relevant initial conditions is therefore needed.

§ *An objection considered*. Wayne Riggs (2007) presents the following case against Pritchard's relativization of Modal Chance 1 to initial conditions:

> Basketball Player
>
> Imagine a young basketball player who has tremendous natural and developed skills and displays them proficiently every day in practice. Unfortunately, he also has a terrible fear of failure that causes him to "choke" when he is actually playing a game against an opposing team. A typical performance during a game would be for him to take, say, 20 shots and miss all 20. And this is due to his fear in-

> terfering with his ability to deploy his impressive skills. But every once in a while, for no reason the player has ever been able to determine, he finds himself confident and calm for a moment or two during a game. One night he happens to have the ball in his hands when this occurs, and he shoots the ball. Absent his usually crippling fear, he makes a skilful shot, which goes in. (Riggs 2007: 339)

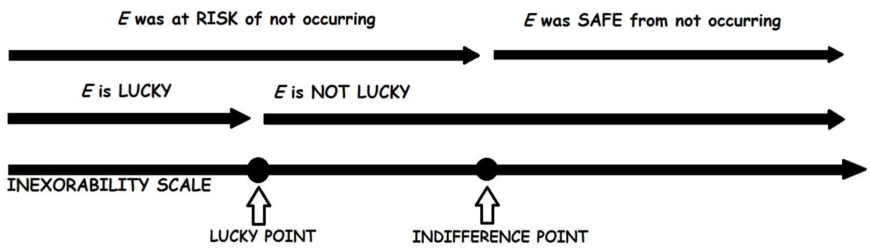
Riggs argues that there is no non-arbitrary way of determining the relevant initial conditions. According to him, if we include the typical fearfulness as part of the initial conditions, we will consider the shot lucky. If we remove that factor, we will conclude that it is not. Riggs thinks that the shot is clearly not by luck because our judgment about the luckiness of the shot is conditioned by our judgment that the player had control over it.

While I agree that our judgments about the luckiness of events are conditioned by our judgments of whether we have control over them, I do not think that it so obvious that the player has control over the shot, since he cannot control his fear. Therefore, it is not obvious that there is no risk of missing the shot and that the case is not a case of luck.

At any rate and for the purposes here, we can use the initial conditions clause to explain the instability of intuitions: we cannot clearly say that the player was at risk of missing the shot or that it is by luck that he has not missed because we cannot specify the relevant initial conditions for the shot. In particular, we do not know whether to hold constant the period of calm and confidence across the relevant close possible worlds or to allow that in some of these worlds the player is not confident enough, and the reason we cannot choose one of these two options is that we need more information on the psychology of the player: under what conditions can the player overcome his fear and under what conditions he cannot? Basketball Player is under-described.

C. In what proportion of close possible worlds should an event fail to occur in order for its actual occurrence to be by luck? The idea, so far, is that lucky events are events that occur in the actual world but that were at risk of not occurring. Our interpretation of the notion of safety (as a point in the inexorability scale that is beyond the indifference point) entails that there can be events that are *at little risk* of not occurring but that are not completely safe from not occurring because it is not necessary that they will occur (in the local sense of necessity determined by context). Suppose that one of these events at little risk of not occurring occurs. Is it by luck that it has occurred? Not necessarily: that a (significant) event was at risk of not occurring does not necessarily makes the occurrence of that event lucky. That is, there is not one-to-one correspondence between lucky events and events that were at risk of not occurring and non-lucky events and events that were safe from not occurring.

We can put these ideas in terms of possible worlds: in order for an event to be considered lucky there must be a large enough proportion of close possible worlds in which the event would *not* occur. We can visually represent it as follows:



As before, the *indifference point* is the point beyond which an event is (was) safe from not occurring, i.e., the point beyond which the actual occurrence of the event is inexorable (the indifference point changes position from context to context). On the other hand, the *lucky point* represents the minimal proportion of close possible worlds in which the event must not occur to be lucky in the actual world. In other words, it represents the maximum degree of necessity or inexorability that a lucky event can have (which can be higher or lower depending on the context). Beyond that point, the actual occurrence of the event is not by luck (although we may still say that the event was at (some) risk of not occurring). That is to say, the indifference point and the lucky point do not necessarily coincide (although they might). This indicates that lucky events are a special type of risky events.

The question is: in what proportion of close possible worlds should an event fail to occur in order for its actual occurrence to be by luck? It might be thought that the proportion of close possible worlds must be a large proportion. This is true of lottery wins, for example. When one wins a lottery in the actual world, in *most* close possible worlds where one buys the same ticket one does not win the lottery.

However, other events can be lucky even though they do not occur in a large proportion of close possible worlds. Consider, for instance, a coin toss. Since the probability of heads is 0.5, we can suppose that in approximately half the close possible worlds the outcome would be still heads and that in the other half the outcome would be tails. The outcomes of coin tosses are intuitively lucky for people. Consequently, it would be too strong to require that, in order for an event to be lucky, the event must not occur in most close possible worlds.[33] A weaker

33 One could think that MODAL CHANCE 1 implies such requirement. However, Pritchard is aware of the problem. He suggests that "the correct reading of 'wide

clause like the one in Coffman's MODAL CHANCE 2 accounts both for the coin toss and the lottery win, as it only requires that, to be lucky, the relevant event must not occur in at least half the close possible worlds. Nevertheless, as Levy (2011: 17-18) correctly suggests, we can accept that if an event that does not occur in half the close possible worlds is lucky, an event that does not occur in little less than half the close possible worlds (e.g., in 49,5% of them) can be lucky as well.

According to Levy, there is no fixed proportion of close possible worlds where an event must not occur to be considered lucky in the actual world. His point is that the threshold varies from case to case as a matter of the significance of the event (this is how the 'large enough' of MODAL CHANCE 3 should be read). This idea, translated in our terms, means that the lucky point of the inexorability scale varies from case to case as a function of the significance of the event.

A similar point, I think, can be made about risky/safe occurrence and, in turn, about the notion of *danger*: the significance of the relevant event helps to determine the position of the indifference point and this affects whether or not we ascribe danger to an event. I do not claim that significance is the only pragmatic factor that determines the position of the indifference point, but it is certainly an important factor.

In the case of luck, we can make the account more specific: the more significant an event is for an agent, the smaller needs to be the proportion of close possible worlds in which it would not occur to be lucky for the agent. In a similar way, the more significant an event is for an agent, the smaller needs to be the proportion of close possible worlds in which it would occur to be considered dangerous for the agent. Let us consider three cases in which someone, *S*, plays Russian roulette:

CASE-1: *S* plays with a six-chambered revolver, three bullets and 0.05 probability that the gun jams (because it is not very clean). We can suppose that in approximately 49,5% of close possible worlds where *S* pulls the trigger she shots herself in the head. Suppose that *S* pulls the trigger but does not get shot. Intuitively, it is by luck that *S* lives. Intuitively, *S* is in clear danger of getting shot.

CASE-2: *S* plays with a twelve-chambered revolver, one bullet and zero probability that the gun jams (because it has been thoroughly cleaned). Accordingly, the probability that *S* dies if she pulls the trigger is approximately 0.08. Suppose that *S* pulls the trigger and does not get shot. It is by luck that *S* is alive? Well,

class' in (L1) [MODAL CHANCE 1] is at least approaching half of the relevant nearby possible worlds, and that typically events which are *clearly* lucky will be events which do not obtain in most nearby possible worlds" (Pritchard 2005: 130). Nevertheless, the outcome of a coin toss can be *clearly* lucky for someone even though it does not obtain in most nearby possible worlds.

surviving in this case is not as lucky as in case 1 but still her survival seems to count as lucky and, similarly, it seems that *S* is still in danger.

CASE-3: *S*'s plays with a water gun. As in case 2, the probability that the gun shoots is approximately 0.08 (the batteries of the gun are running low so that most of the times the gun does not have enough power to shoot water). Suppose that *S* pulls the trigger and the gun does not work. Is *S* lucky for not getting squirted by a stream of water? Intuitively, if *S* is lucky, she is clearly not as lucky as in case 2. In the same way, if *S* is in danger, she is clearly not in so much danger as in case 2.

What makes the difference between the cases? While the difference between cases 1 and 2 is just a matter of modal robustness (in case 1, *S* survives in less close possible worlds than in case 2), the difference between case 2 and 3 is not: the probability of the events in question is equally low (so that we can suppose that they would obtain in the same number of close possible worlds). The difference is that in case 3 the relevant event (getting wet) is much less significant for *S* than in case 2 (dying). In general, the less significant an event is, in more close possible worlds should the event not occur to be considered lucky. In addition, the less significant an event is, in more close possible worlds should the event occur to be considered dangerous. Accordingly, in case 2 *S* is in danger of dying and her subsequent survival counts as lucky even though *S* would survive in most close possible worlds: the event is very significant for *S*. By contrast, *S* is not (very) lucky to or not (much) in danger of remaining dry even though *S* would get wet in the same proportion of close possible worlds in which *S*, in case 2, dies. The reason is simple: getting wet is not as significant as dying.[34]

§ A modal chance condition for luck. The two general points that we have learned from the discussion above are that 1) we need a clause on relevant initial conditions (because it does explanatory work) and that 2) the proportion of close possible worlds in which an event should fail to occur in order for its actual occurrence to count as lucky for an agent can vary from case to case as a function of the significance that the event has for the agent. Keeping these two points in mind, we can formulate the following modal chance condition for luck:

- MODAL CHANCE: *E* is lucky for *S* only if *E* occurs in the actual world but would not occur in a large enough proportion of close

34 Let me answer the eventual question of *how* significance determines the relevant proportion of possible worlds in each context with a quote by Levy: "It would be a mistake to seek much greater clarity about how significance affects how large a proportion of nearby worlds must deviate from the actual world in order for the event to count as lucky: luck is a vague concept, and an adequate account of it must be vague (in the same way and in the same places) as well" (Levy 2011: 18).

> possible worlds where the relevant initial conditions for E are the same as in the actual world.[35]

Let us recap. We have found two possible ways of conceiving the event-focused sense of risk: Objective Modal Risk (E is at risk of occurring at t just in case E would occur at t in a large enough proportion of close possible worlds) and Objective Probabilistic Risk (E is at risk of occurring at t if and only if there is high probability that E will occur at t). Both are legitimate and both capture different uses of the term 'risk' (the use of 'risk' in scientific and technical contexts tends to be probabilistic; in ordinary contexts, modal). Both are useful to specify the idea that lucky events are events that obtain but that were at risk of not obtaining. In particular, the probabilistic conception of risk leads to two chance conditions for luck in terms of objective probability: Objective Probabilistic Chance (E is lucky for S at t only if, prior to the E's occurrence at t, there was low objective probability that E would occur at t) and Conditional Probabilistic Chance (E is lucky for S at t only if, prior to E's occurrence at t, there was low objective probability conditional on C that E would occur at t). The modal conception of risk leads to several chance conditions for luck in terms of modality. Modal Chance is my preferred one.

The idea that luck is a form of risk seems right. The question now is whether the relevant sense of risk should be the probabilistic or the modal one. That is, is the nature of luck modal or probabilistic? In most cases, it makes no difference whether we explain the cases in probabilistic or in modal terms. For example, a lottery win is a lucky event that obtains in the actual world but that was at probabilistic/modal risk of not obtaining. Almost any case of luck can be indistinctively explained in modal or in probabilistic terms. However, I will show next that there are some events that are intuitively lucky but that had a high (conditional or unconditional) probability of occurring, which means that probabilistic chance conditions are not necessary for luck. This kind of events can be explained in modal terms, which means that, although most cases of luck can be explained in probabilistic terms, the ultimate nature of luck is modal, not probabilistic. To reach this surprising conclusion we must follow the lead of Timothy Williamson for a while.

35 The reason I do not include indexes to specific times is that, in certain cases, luck is due to the fact that the target event could have occurred at a time slightly different to the time at which it occurs in the actual world. For example, in Russell's stopped clock case an agent could easily have formed a false belief about the time because she could easily have looked at a stopped clock at a time different to the time at which she looks at it in the actual world (Russell 1948: 170-171).

1.6.5 *Highly Probable Lucky Events*

Williamson (2009) identifies an important divergence between the probabilistic and the modal conceptions of risk: there are some cases in which the modal conception implies that there is no risk that an event will occur despite the high probability that the event has of occurring.

How will this divergence affect chance conditions based on OBJECTIVE MODAL RISK? If OBJECTIVE MODAL RISK entails that some highly probable events were at risk of not occurring, then modal chance conditions formulated on this basis (e.g., MODAL CHANCE) will entail that it is by luck that those events have occurred. However, how can it be that it is by luck that an event occurs if the event had high probability of occurring? Contrary to what one may understandably think, this puzzle does not undermine the view that lucky events are modally risky events. Rather, it ratifies it.

To see exactly where OBJECTIVE MODAL RISK and OBJECTIVE PROBABILISTIC RISK diverge, we need to present some of the arguments that Williamson provides concerning the ordinary notion of safety. According to Williamson, there are two ways of conceiving safety: there is a 'small risk' conception and a 'no risk' conception. On the 'no risk' conception, *S* is safe from *X* if only if there is no risk that *X* will obtain. By contrast, on the 'small risk' conception *S* is safe from *X* if only if there is small risk that *X* will obtain. The 'no risk' conception is the adequate one, Williamson (2009: 11-12) argues, because it allows us to guarantee the validity of arguments like the following:

S was safe from being shot by *X*.
S was safe from being shot by *Y*.
S was safe from being shot by *Z*.
S was safe from being shot other than by *X*, *Y* or *Z*.

S was safe from being shot.

Williamson's idea is that if we substitute the term 'safe' with the expression 'at no risk', the argument is still valid:

S was at no risk of being shot by *X*.
S was at no risk of being shot by *Y*.
S was at no risk of being shot by *Z*.
S was at no risk of being shot other than by *X*, *Y* or *Z*.

S was at no risk of being shot.

However, the following substitution does not yield a valid argument:

S's risk of being shot by *X* was small.
S's risk of being shot by *Y* was small.

S's risk of being shot by *Z* was small.
S's risk of being shot other than by *X*, *Y* or *Z* was small.

S's risk of being shot was small.

Therefore, the ordinary notion of safety should be understood along the parameters of the 'no risk' conception (this is consistent with Williamson's idea that safety is a sort of local necessity). Nevertheless, Williamson identifies a possible problem concerning this way of understanding safety. If an event has extremely low probability of occurring, and the event has still not occurred, we can plausibly consider that there is no risk that the event will occur (as it follows from OBJECTIVE PROBABILISTIC RISK). However, if the number of extremely low probable events of that type is sufficiently large at a certain time, the probability of their disjunction is high, that is, the probability that at least one event of that type occurs is high.[36] The puzzle arises when, by an argument like the ones presented above, we conclude that there is no risk that at least an event of that type will occur. To illustrate this, Williamson uses examples of bizarre 'quantum' events like the following:

> TUNNELING
>
> Consider [the] extremely unlikely (...) event that a marble I drop tunnels through the whole house and lands on the ground underneath, leaving the matter it penetrates intact. On natural interpretations according to which the wave function represents facts of objective chance, such events are not merely nomologically possible, but have a non-zero chance of occurring. When I drop a marble, the situation can be redescribed as a cosmic lottery with immensely many tickets. In this lottery, holding a winning ticket means having one's marble tunnel through the house. (Hawthorne and Lasonen-Aarnio 2009: 94)

The idea is that if we apply the second argument above to a sufficient number of instances of marble dropping, such that for each instance we can claim truly that there is no risk that the marble tunnels through the whole house, we can end up concluding that there is no risk that a marble tunnels through the whole house, despite the high probability that at least one instance of marble dropping will have that result. This kind of examples are not very intuitive. To grasp better the puzzle, I propose the following example, which will also allow to reach the desired conclusion that there are highly probable lucky events:

36 The probability of a disjunction of independent events (such as getting a job at different places) is: $\Pr(P_1 \vee P_2 \vee ...P_n) = 1 - (\Pr(\neg P_2) \cdot (\Pr(\neg P_1) \cdot (\Pr(\neg P_n)$. Thus, the higher n, the more probable the disjunction.

Lazy Luke

> In a distant future, the galaxy is populated by billions and billions of people. The billions of corporations of the Galactic Empire are hiring computer technicians. Luke is an unemployed computer technician, and very lazy. He does not want to work. All he wants is to lie down on the couch and play videogames. However, the Galactic Empire's political system forces unemployed people to apply for jobs constantly, so Luke reluctantly switches on his supercomputer and starts applying for billions of jobs. Luke, who is after all a clever guy, makes sure to upload a very bad CV. In fact, he makes sure to upload the worst CV of the galaxy (he knows how to do that). Hiring decisions are made based on the number of candidates and the quality of their CV's, so by submitting a disastrous CV, Luke makes sure that whenever there is another candidate, he will not be chosen. Furthermore, he knows that his name is on the I.D.L.E list, i.e., the list of Individuals Devoted to Leisure and Enjoyment, which contains the names of those who should never be hired because of their extreme laziness (all companies use I.D.L.E). The job competition is tough, so for every single job there are millions of applications. In addition, people normally inflate their CV's.

For any n, the probability that Luke will get job J_n is extremely low (he submits the worst CV, there are millions of applicants for each job offer, he is on I.D.L.E, and so on). Thus, it seems that, for any n, there is *no risk* that Luke will get J_n, because there are always many better prepared applicants for the job.

Suppose that the number of applicants for each job offer, the bad quality of Luke's CV and the fact that his name is on I.D.L.E is stable in close possible worlds. That is, possible worlds in which Luke intentionally uploads an excellent CV or in which he is the only applicant or in which his name is not on I.D.L.E are distant. This being so, there is no reason why Luke should make himself safe from these possibilities.

Nevertheless, it seems that Luke should be worried about the high probability of getting *at least* one job. The probability that Luke will get at least one job is the probability of the disjunction of the following propositions: <Luke will get J_1>, <Luke will get J_2 >, ...<Luke will get J_n>. Although the probability that Luke gets a job at one particular corp is extremely low, if Luke applies for a very high number of jobs (as he does), the probability that he gets at least on job is high (i.e., the probability of the disjunction is high).

Now, suppose for a moment that, in the end, Luke remains unemployed despite the high probability of getting at least one job. By an argument like the arguments presented before we can conclude that Luke was at no risk of getting a job.

Luke was at no risk of getting J_1
Luke was at no risk of getting J_2
...Luke was at no risk of getting J_n
Luke was at no risk of getting a job other than $J_1 ... J_n$

Luke was at no risk of getting a job.

However, this conclusion strikes us as wrong. How can it be that Luke was at no close risk of getting a job if the probability the he would get at least one job was extremely high? Williamson thinks that the consequence of counting as safe an event that has high chance of occurring is the price that we must pay for a practically tractable conception of safety.

Nevertheless, contrary to what may initially seem, this is not a bad consequence. Note that when our hero applies for jobs, he makes sure that *for each job offer* there are more applicants with a better CV and that his name is on I.D.L.E. By doing so, he makes himself safe from the highly probable eventuality of getting at least one job. According to Williamson, one virtue of the 'no close risk' conception of safety is precisely that it "permits us to make ourselves safe from a disjunction of dangers by making ourselves safe from each disjunct separately, and to check that we are safe from the disjunction by checking that we are safe from each disjunct in turn" (Williamson 2009: 17). Let us put an end to the Williamsonian story and let us return to the question of whether the nature of luck is modal or probabilistic.

§ *A counterexample to objective probabilistic chance conditions for luck*. The result that Luke is at no risk of getting a job may be seen as problematic for chance conditions based on OBJECTIVE MODAL RISK, because if there is no risk that Luke will get a job, they will entail that the fact that he does not get a job is not by luck. Again, how can it possibly be that Luke is not lucky to remain unemployed if he had every chance of getting at least one job?

No one denies that if a highly probable event (e.g., getting at least one job) unexpectedly fails to occur, in most cases it will be by luck that it does not occur. However, it is *not necessarily true* that if an event has high probability of occurring and fails to occur, it is by luck that it does not occur. Sometimes, as LAZY LUKE proves, the nonoccurrence of a highly probable event is not by luck. I will turn LAZY LUKE into a clear case of luck in a moment, but first note that from the fact that it is highly probable that Luke gets at least one job (i.e., from the fact that the relevant disjunction has high probability), it does not follow that in *close* possible worlds where he applies for all job offers, he gets at least one job. The reason is that, as we have said, for every single job offer Luke makes sure that he submits the worst CV, that there are more applicants and that he is on I.D.L.E. In this way, he makes himself safe from the eventuality of getting a job. This means that, *for each application*, it is *not by luck* that he remains unemployed.

From the fact that Luke makes himself safe from the eventuality of getting a job it does not follow that in no possible world he does not get a job. There are, for sure, possible worlds in which he does get a job. However, those worlds are too distant to make us judge that he is at risk of getting a job. Those distant possible worlds are, for example, worlds in which the one million applicants for certain job die unexpectedly and the corporation in question has no option but to hire Luke (because, say, the law obliges the corp to fill the position). In brief, what follows from the fact that Luke makes himself safe from the eventuality of getting a job is that in *close* possible worlds he does not get a job. Consider now the following possible situation:

> Job
>
> Luke makes sure to upload the worst CV for each job offer, to check that there are millions of applicants, to check that he is on I.D.L.E, and so on. Unbeknownst to Luke, there is a problem with the application sent to Microcorp. Due to some unusual interference in the data stream, the contents of the CV that he has sent change in such a way that the human resource department of Microcorp receives a CV full of so many brilliant achievements that they decide to hire Luke instantly (by law, once a corp hires a worker, the worker cannot be fired for a period of one year).

Intuitively, it is by (bad) luck that Luke gets the job, so it is by luck that he gets at least one job. Objective Probabilistic Chance says that an event (e.g., getting a job) is lucky for an agent only if, prior to the occurrence of the event, there was low objective probability that the event would occur. As we have seen, the probability of getting at least one job was very high. Therefore, low probability of occurrence cannot be necessary for luck.

The result is the same in the case of conditional probabilities. Recall Conditional Probabilistic Chance: E is lucky for S at t only if, prior to E's occurrence at t, there was low objective probability conditional on C that E would occur at t, where C is some condition to be specified. In general, for any job J_n, C might be the fact that Luke makes sure to upload the worst CV for J_n, the fact that he checks that there is a considerable number of applicants for the job, the fact that he checks that he is on I.D.L.E, and so on. As regards the job at Microcorp, C could also include the fact that there is an interference in the data stream that changes the contents of Luke's CV in such a way that the corp receives a CV full of brilliant achievements. The probability that Luke gets a job at Microcorp conditional on C so understood is certainly very low. But, once again, for a sufficient number of conditionalized disjuncts the probability that he gets at least one job is very high. Yet it is by luck that Luke gets a job at Microcorp and, therefore, it is by luck that he gets at least one job.

§ *The modal nature of luck*. MODAL CHANCE gives an intuitive explanation of JOB: although the probability that Luke gets at least one job is high, it is by luck that Luke gets a job because in most close possible worlds in which he applies for the job offer at Microcorp there is no data interference and he does not finally get the job. In addition, in most close possible worlds he does not get a job in any other corp. To see this, let me compare LAZY LUKE with a lottery case.

In a lottery case, there is high probability that at least one ticket T will win (since the number of tickets is large, the disjunction of the following propositions is highly probable: <T_1 will be the winner>, <T_2 will be the winner>, ...<T_n will be the winner>).[37] Analogously, the probability that Luke will get at least one job is also very high. In addition, in the lottery case, the probability that a particular ticket will win is extremely low. Analogously, the probability that Luke will get a particular job is extremely low.

However, there is an important difference between them. If one buys all the tickets of a lottery, given the high probability that at least one ticket will be the winner, there is close risk that one will win the lottery. However, why is it that there is no close risk that Luke will be hired given that he has applied for all job offers and hence the probability that he will get at least on job is very high? The difference is that, for each job offer, Luke makes himself safe from the eventuality of getting a job (he submits the worst CV, he checks that there are more candidates, that he his on I.D.L.E, and so on). Luke's *actual* actions prevent possible worlds in which he would get a job from being close, because close possible worlds are similar to the actual world and worlds in which he gets a job are worlds in which he does not perform such actions. To compare: lottery players are typically not in a position to make themselves safe from the eventuality of winning, no matter the number of tickets they have, and for this reason close possible worlds in which they win are close. They can make themselves safe from the eventuality of winning, for example, by manipulating the lottery system. In this way, worlds in which they win would not be close.

1.7 LUCK AND FORTUNE

In this next section, I will distinguish between two senses of luck (one of which will receive the name 'fortune'). I will use a counterexample

37 The probability is very high but is not 1 because there is a very low probability that there is no winner at all (e.g., if the lottery device breaks down).

by Jennifer Lackey to the necessity for luck of Pritchard's MODAL CHANCE 1 (and by extension a counterexample to the necessity for luck of any modal chance condition) to make the distinction. In turn, the distinction will help us to explain Lackey's counterexample.

1.7.1 *Lackey Against Modal Chance Conditions*

Jennifer Lackey (2006, 2008) thinks that the following case shows that MODAL CHANCE 1 is not necessary for luck:

> BURIED TREASURE
>
> Sophie, knowing that she had very little time left to live, wanted to bury on the island she inhabited a chest filled with all of her earthly treasures. As she walked around trying to determine the best site for proper burial, her central criteria were, first, that a suitable location must be on the northwest corner of the island, where she had spent many of her fondest moments in life, and secondly, that it had to be a spot where rose bushes could flourish, since these were her favourite flowers. As it happened, there was only one particular patch of land on the northwest corner of the island where the soil was rich enough for roses to thrive. Sophie, being excellent at detecting such soil, immediately located this patch of land and buried her treasure, along with seeds for future roses to bloom, in the one and only spot that fulfilled her two criteria. One month later, Vincent, a distant neighbour of Sophie's, was driving in the northwest corner of the island, which was also his most beloved place to visit, and was looking for a place to plant a rose bush in memory of his mother who had died ten years earlier, since these were her favourite flowers. Being excellent at detecting the proper soil for rose bushes to thrive in, he immediately located the same patch of land as Sophie had found one month earlier. As he began digging a hole for the bush, he was astonished to discover a buried treasure in the ground. (Lackey 2006: 285)

According to Lackey, BURIED TREASURE is a counterexample to MODAL CHANCE 1 because, while Vincent's discovery is intuitively by luck, it would occur in most close possible (MODAL CHANCE 1 says that lucky events would not obtain in most close possible worlds).[38] By contrast, it is not so obvious that BURIED TREASURE is a counterexample to MODAL CHANCE and MODAL CHANCE 3. Since they say that lucky events would not occur in a *large enough* proportion of close possible worlds, it could be argued that in BURIED TREASURE the proportion of

38 *A fortiori*, BURIED TREASURE is also a counterexample to MODAL CHANCE 2: Vincent finds the treasure in more than half the close possible worlds.

close possible worlds in which Vincent would not find the treasure is large enough to consider his actual discovery lucky. However, according to Modal Chance and Modal Chance 3 the size of the proportion of close possible worlds in which an event should not obtain in order to be considered lucky for an agent is fixed by the significance of the event for the agent. Typical situations in which we consider lucky to find a treasure are situations in which one would find the treasure in *most* close possible worlds. How can it be explained the difference between Vincent's discovery and typical discoveries of treasures if the significance of the discoveries is always the same? Buried Treasure is problematic for *any* modal chance condition for luck.

No philosophical theorizing will be able to annul the intuition that the case is a case of luck, but philosophical theorizing can certainly help to neutralize Lackey's point that modal chance conditions are not necessary for luck. My aim in what follows is to show that Lackey is wrong in thinking that modal chance conditions are unnecessary while conceding her that the case is a case of luck. To do that, a distinction between two types of luck will be needed, but first let me explain why Lackey thinks that the case is a case of luck.

§ Luck out of coincidence? For Lackey, the reason the case is a case of luck is the extraordinary coincidence that Vincent digs in the exact place where Sophie buried the treasure. Lackey claims: "it is precisely because of this fortuitous combination of circumstances that the discovery of the buried treasure is so clearly a lucky event" (Lackey 2008: 263). However, is this coincidence the reason Vincent's discovery is lucky?

Let us show that the reason Vincent's discovery is lucky is not such a "fortuitous combination of circumstances". Remember Owens's account of coincidences (Owens 1992): a coincidence is an inexplicable combination of events; it is inexplicable because its constituents are produced by *independent causal factors* in such a way that we cannot explain why those constituents come together. In Buried Treasure, one constituent is the fact that Sophie's stable dispositions make the treasure be located in that specific place (constituent 1), the other constituent is the fact that Vincent's disposition to detect proper soil makes him dig in that specific place (constituent 2). My contention is that, although such a coincidence certainly involves luck (all coincidences involve luck), the reason Vincent's discovery is lucky is not that coincidence.

To see this, suppose that constituent 1 is rather the following: Fortuna (the goddess of fortune) knows that Vincent is excellent at detecting proper soil for rose bushes and that he wants to plant a rose bush in memory of his mother; she also knows that the only patch of land on the island where the soil is rich enough for roses to thrive is *L* and she knows that Vincent knows that. Fortuna buries a treasure at *L* (she knows that Vincent will dig there). Vincent goes to *L*, digs and finds the treasure. "How lucky I am!", he thinks. Fortuna smiles.

Once the case is described in this way, Vincent's discovery is no longer a coincidence. Yet, it is clearly by luck. The case so described is structurally equivalent to Lackey's (Vincent's discovery is still modally robust), which indicates that the reason why Vincent's discovery is by luck is not a "fortuitous combination of events". As we will see, Vincent's discovery is by luck because he lacks control over the discovery, but first let us analyze Pritchard and Levy's replies to Lackey's objection.

§ Pritchard and Levy's replies. For Pritchard (2005: 144, fn. 15), it is not correct to say that an event that occurs in the actual world is lucky for an agent if the event would still occur in close possible worlds. For example, he thinks that it is not correct to say that an event is lucky for an agent if the event would occur in most close possible worlds due to the secret intervention of a conspirator (e.g., the goddess Fortuna). Pritchard thinks that it is not correct to say so, because if the agent affected by the event in question discovered that everything had been carefully planned all along, she would plausibly no longer regard it as a lucky event.

Lackey (2008: 262-263) replies that in Buried Treasure there is no deliberate intervention of any sort. Rather, the modal robustness of Vincent's discovery is due to a fortuitous combination of circumstances (as said, the causal independence between Sophie and Vincent's dispositions is crucial for Lackey) in such a way that Vincent would still think that he has been lucky to have found the treasure even after hearing all the details surrounding the discovery.

Levy (2009: 494) tries to give further support to Pritchard's point by presenting a case like Buried Treasure in which a conspirator plans everything all along (a case similar to my modification of the case). In that modified case, Levy argues, Vincent's discovery is not by luck and then he claims:

> Of course in the original *Buried Treasure*, there is no one ensuring that Vincent finds the treasure. Instead, people's stable dispositions play this role. Because these dispositions are stable, they are held fixed across nearby possible worlds, like the plans of [Fortuna]. (Levy 2009: 494)

Levy argues that if Vincent were appraised of all the facts, he might not withdraw the claim that he is lucky. However, according to Levy, if he came to know the relevant similarities between the modified case and Buried Treasure (in particular, if he realized that he could not easily have failed to find the treasure) he certainly *should* withdraw that claim.

§ Is there an ordinary distinction between luck and fortune? Pritchard and Levy's replies do not seem very effective against Lackey's intuition

that in BURIED TREASURE it is by luck that Vincent finds the treasure. Vincent's discovery certainly seems by luck. Perhaps because that intuition is so strong, Pritchard and Levy (also Coffman 2007) adopt a further strategy that aims to defend the necessity of modal chance conditions for luck while explaining why we have a sense of 'luckiness' in cases of this sort. Their claim is that this kind of cases are not cases of luck but of *fortune*. How do they understand the notion of fortune?

For Pritchard (2005: 144, fn. 15), fortunate events are events that count in one's favor over which one has no control (consequently, according to Pritchard, positively lucky events are part of a more general class of fortunate events). For Levy (2009: 495-496), fortunate events are non-lucky events that have luck in their causal history (namely, in their proximate causes). He explains that luck in the circumstances (e.g., the coincidence that Sophie buries the treasure at the same place where Vincent digs to plant a rose bush) is not inherited by the actions performed in those circumstances or by the events that result from them (Vincent's discovery). Finally, in order to characterize the notion of fortune Coffman (2007: 392) appeals to the intuition that, while a lottery winner enjoys a stroke of good luck, a lottery loser does not suffer a stroke of bad luck, provided that it was very likely that she would lose. A lottery loser is, according to Coffman, merely unfortunate.

If Pritchard, Levy and Coffman's distinctions are distinctions of ordinary concepts, it follows that on certain occasions competent speakers of English misapply the terms 'luck', 'lucky', 'fortune' and 'fortunate'. For example, for Pritchard, it would be an error to say that it is bad fortune that I lost the lottery (for him, fortune only refers to events that count in one's favor). According to Levy, it would be incorrect to say that I am fortunate to have won a fair lottery (for him, fortunate events are not chancy). Similarly, for Coffman it would be a mistake to claim that it is bad luck that I have lost the lottery.

However, do we really speak falsely when we make such claims? It does not seem so. The terms 'lucky' and 'fortunate' (and 'luck' and 'fortune') can be interchangeably used in ordinary discourse without risk of falsity or infelicitousness. Furthermore, Rescher (1995) has intuitions about these terms that are opposite to those of Pritchard, Levy and Coffman. According to Rescher, "you are fortunate if something good happens to or for you in the natural course of things. But you are lucky when such a benefit comes to you *despite its being chancy*" (Rescher 1995: 28; emphasis mine). All these considerations put together indicate that, from an ordinary point of view, there is no such a distinction between luck and fortune: the terms 'luck' and 'fortune' refer to the same ordinary concept.

§ *A technical distinction*. Since competent speakers use the terms 'luck' and 'fortune' and 'lucky' and 'fortunate' indistinctively, the dispute between Lackey and defenders of modal chance conditions might seem merely terminological. The most charitable way of interpreting Pritchard, Levy, Coffman and also Rescher's efforts is as if they were introducing technical distinctions that serve some philosophical purpose.

As I see it, what these authors are trying to distinguish are two different phenomena that our ordinary usage of the terms 'lucky' and 'fortunate' overlook. In ordinary discourse we apply the terms 'lucky' and 'fortunate' to a person who has won a fair lottery, but also to a person who has won a lottery which, unbeknownst to her, someone rigged in her favor. However, winning a fair lottery does not seem to have the same quality of 'luckiness' or 'fortunateness' as winning a lottery rigged in one's favor. The two different intuitions that these cases elicit points to the existence of two different senses of the notion to which the terms 'luck' and 'fortune' refer in ordinary discourse.

1.7.2 *Risk, Luck and Fortune*

1.7.2.1 *Four Combinations of Risks*

In my view, the difference between the two senses of luck/fortune is a difference in risk. To see this, we need to recall the distinction between the two senses of risk. On the one hand, an *event* can be at risk of occurring (or of not occurring). I called this the *event-focused* sense of risk. On the other hand, an *agent* can be at risk with respect to an event. I called this the *agent-focused* sense of risk. Recall that the event-focused sense of risk has been understood in modal terms.[39] The agent-focused sense of risk, on the other hand, has been understood in terms of lack of control.[40] By combining these two senses of risk, we come up with four different structures of cases. Cases in which:

A. *S* is at risk with respect to *E* & *E* is at risk of occurring.

B. *S* is at risk with respect to *E* & *E* is not at risk of occurring.

C. *S* is not at risk with respect to *E* & *E* is at risk of occurring.

D. *S* is not at risk with respect to *E* & *E* is not at risk of occurring.[41]

A-*cases*. They constitute paradigmatic cases of luck (or fortune). A good example of an A-case is a fair lottery win. When one wins a fair lottery, prior to winning, *losing the lottery* was at big risk of happening, i.e., one would lose in most close possible worlds (event-focused sense of risk). In addition, if the lottery is fair, *one* is at risk of losing

39 Objective Modal Risk*:* *E* is at risk of occurring at *t* if and only if *E* would occur at *t* in a large enough proportion of close possible worlds.
40 Agential Risk*:* *S* is at risk with respect to *E* if and only if *S* lacks control over *E*.
41 Depending on the event and the context in question the kind of risk at issue can be, instead of risk of occurrence, risk of nonoccurrence.

because one has no control over the lottery process and its outcomes (agent-focused sense of risk).[42]

B-*cases*. They also constitute cases of luck (or fortune). A B-case would be, for example, a case in which one wins a lottery because, unbeknownst to one, the organizer has rigged the lottery in one's favor. In this case, there was no risk of *losing the lottery*, because the organizer takes care of manipulating the lottery system in such a way that one wins in the actual and in all close possible worlds (event-focused sense of risk). Still, *one* is at risk with respect to the outcome of the lottery because one has no control over the lottery process (agent-focused sense of risk). For this reason, we can intuitively say that one is lucky (or fortunate) to win.

D-*cases*. In D-cases, by contrast, there is no risk whatsoever, so we should not expect the presence of luck (or fortune) in them. A D-case would be, for example, a case in which one rigs a lottery in favor of oneself. Winning in that case would not be by luck because 1) the eventuality of *losing the lottery* is at no risk of happening (event-focused sense of risk) and 2) *one* was never at risk of losing, since one has control over the lottery process (agent-focused sense of risk).

C-*cases*. In this kind of cases, the relevant *event* is at risk of occurring (or of not occurring) and yet *one* is at no risk with respect to the event. Excellent examples of C-cases are decisions made on a whim (*whimsical decisions*). Suppose that one decides to go to Paris for the weekend on a whim. Since one has made the decision on a whim, the decision was at big risk of not being made (event-focused sense of risk). However, *one* is at no risk of not deciding to go to Paris because, even though the decision was made on a whim and one could easily have not made it, it was a self-consciously made decision after all, which means that one had control over the (precipitated) deliberation process (agent-focused sense of risk). Interestingly, even though one could easily have not made the decision, it is not by luck that one makes it.[43] In this sense, C-cases are not cases of luck (or fortune).

Lackey (2008) thinks that *whimsical events* (events that result from decisions made on a whim) prove that MODAL CHANCE 1 plus a sig-

42 If one does not have control over the lottery process and its outcomes, *one* is also at risk of winning (remember that the notion of risk that we are using can be also applied to positive events). In general, we are attracted by lotteries because they allow us to expose us to the whims of luck intentionally by engaging in a game whose relevant parameters are beyond our control. This lack of control gives us hope of winning even though we know that the probability of losing is extremely high.

43 The same applies to *torn decisions*, which Mark Balaguer (2010) uses to defend a naturalistic libertarian account of free will. He defines torn decisions as the kind of decisions we sometimes make when we have reasons for two or more options and we feel torn as to which reason is the best, so we end up just choosing one of the options (Balaguer 2010: 71).

nificance condition (*E* is lucky for *S* only if *E* is significant for *S*) are not jointly sufficient for luck. However, although I agree with Lackey that they are not jointly sufficient (my view is that a lack of control condition is needed as well), I think that Lackey's reasons for thinking so are wrong.

Lackey thinks that whimsical events are not lucky (I agree with that). Since whimsical events are the kind of events that result from whimsical decisions, and because this kind of decisions are modally fragile, she thinks that if an event that results from a decision made on a whim occurs in the actual world, the event would not occur in close possible worlds. That might be true. However, it does not necessarily follow from that that Modal Chance 1 plus a significance condition are not jointly sufficient for luck. In particular, Lackey's error in thinking that whimsical events prove that conclusion is due to a misconception concerning the clause on initial conditions of Modal Chance 1.

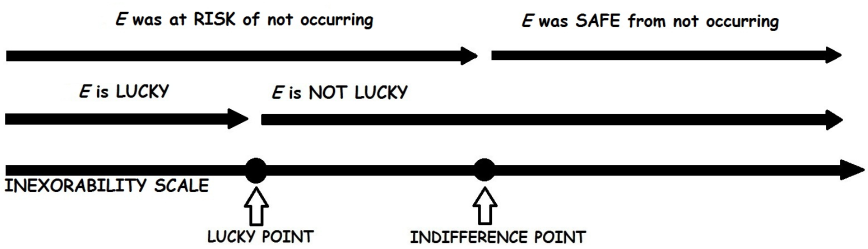
While it might be true that, if one decides to go to Paris on a whim, one would not go to Paris in most close possible worlds, the only close possible worlds that are relevant to assess whether it is by luck that one goes to Paris in the actual world are possible worlds in which the relevant initial conditions for the occurrence of the event are the same as in the actual world. As a general rule, one's decision to φ is always among the relevant initial conditions for one's φ-ing. Therefore, close possible worlds in which the relevant initial conditions for going to Paris are the same as in the actual world are worlds in which one makes the decision to go to Paris. In all close possible worlds in which one decides to go to Paris, one goes to Paris. Consequently, Modal Chance 1 rules out one's going to Paris as a case of luck. Similar considerations apply to Modal Chance. According to both chance conditions, whimsical events are not lucky events.

We can conclude that there is no luck or fortune if there is no risk whatsoever (d-cases) and if there is only risk in the agent-focused sense (c-cases). That is, luck and fortune arise only if the the agent in question lacks control over the relevant event (only if *she* is at risk with respect to the event). As we have seen, there are two types of cases where this occurs: cases where the occurrence or nonoccurrence of the relevant event is modally fragile or, more precisely, where the event in question was at risk of occurring or of not occurring (a-cases) and cases where the occurrence or nonoccurrence of the relevant event is modally robust or, more precisely, where there was no risk that the event occurred or did not occur (b-cases).

Accordingly, one may think that the difference between the two senses of the notion to which the terms 'luck' and 'fortune' refer in ordinary discourse is just the difference existing between a-cases and b-cases, which is, in turn, the difference between risk and safety of occurrence (or of nonoccurrence). However, as we will see next, this is not exactly so.[44]

1.7.2.2 *Two Ways of Being Lucky (or Fortunate)*

Suppose that *E* is an event over which the agent lacks control (i.e., an event with respect to which the agent is at risk). Recall the representation of lucky events in the inexorability scale:



As I explained in section 1.6.4, the lucky point represents the minimal proportion of close possible worlds in which an event must not occur in order for its occurrence in the actual world to be by luck; in other words, it represents the maximum degree of necessity or inexorability that the actual occurrence of an event can have while still counting as lucky. Call an event that captures this sense of 'luckiness' or 'fortunateness' a *lucky*-I or a *fortunate*-I event (an LF-I event for short). The idea is that, beyond the lucky point, an event is no longer LF-I because its actual occurrence is 'too' inexorable.

As we saw, however, the degree of inexorability that prevents an event from being LF-I needs not be the same degree of inexorability that makes an event safe from not occurring. For example, suppose that you are walking to work and you find the only coin on the street. You are certainly LF-I to have found that coin because in most close possible worlds you would not find it (presumably, in those worlds you would be looking at your cell, distracted, and so on). Suppose now that someone has dropped one million coins on the street and you find one of them. You are lucky or fortunate for that, but you are surely not LF-I because in most close possible worlds you would find a coin. You are *lucky*-II or *fortunate*-II (LF-II for short).

The distinction between being LF-I and being LF-II is certainly a technical distinction. However, it is not a mere technicality, as it captures the two aforementioned senses of the notion to which the terms 'lucky' and 'fortunate' refer in ordinary discourse. That is, in ordinary discourse the terms 'lucky' and 'fortunate' indistinctively refer to LF-I and LF-II. In other words, the extension of the predicates 'lucky' and 'fortunate' as ordinarily used is the class of LF-I and LF-II events. For

44 Note that the distinction I am pursuing is not made by other authors.

simplicity, from now on and unless otherwise indicated (here comes an *important terminological point*), I will call the sense in which an event is LF-I *luck* and the sense in which an event is LF-II *fortune*. Keep in mind, however, that these technical terms do not pick out the ordinary notions of luck and fortune, which are the same thing from the layman's point of view.

Let us return to the inexorability scale. The lucky point and not the indifference point is what marks the difference between *luck* and *fortune*. As I argued in section 1.6.4, the location of the lucky point might vary from case to case as a function of the significance of the relevant event (significance might also have influence on the position of the indifference point). For example, for not very significant events like finding a coin, the lucky point may be located at the beginning of the line: that the point is located at the beginning of the line means that the occurrence of the event in the actual world (finding the coin) is not very inexorable, which means, in turn, that the proportion of possible worlds in which one would not find the coin should be considerably large in order for one's actual finding of the coin to count as lucky.

If the lucky point is located near the indifference point, this means that the event could occur in a greater proportion of close possible worlds while still counting as lucky in the actual world. Plausibly, this is explained by the fact that the event is very significant. For example, suppose that you survive a round of Russian roulette with a twelve-chambered revolver, one bullet and zero probability that the gun jams (because it has been thoroughly cleaned). Although the probability that you died was approximately 0.08 and, consequently, there was almost no risk of dying (event-focused sense of risk), you are still lucky to survive (surviving is very significant for you).

Of course, there is a threshold below which your survival would not count as lucky, because the risk of dying was extremely low. For example, suppose that, unbeknownst to you, someone manipulates the revolver in such a way that the probability that you get shot is 0.0001. Since you do not know it, you do not have control over the outcome of pulling the trigger, i.e., *you* are at risk of being shot (agent-focused sense of risk). The proportion of possible worlds in which you would die is not large enough to consider your survival lucky in the technical sense (LF-I). The proportion of worlds exceeds the limit marked by the lucky point. Your survival is, so to speak, 'too' inexorable to be by luck and, therefore, it is by fortune (LF-II).

We can now clearly explain why the distinction between luck and fortune does not correspond to the difference between A and B-cases. First, in an A-case the relevant event is at *risk* of occurring (or of not occurring), whereas in a B-case the relevant event is *safe* from occurring (or from not occurring). In other words, while in an A-case the event in question could have occurred (or not occurred) in close possible worlds, in a B-case there is no close possibility that it would occur (or not occur), i.e., its occurrence or nonoccurrence is necessary in a local sense of necessity (given the relevant context).

Second, a locally necessary event is fortunate for some agent when the agent lacks control over it (e.g., winning a lottery that, unbeknownst to one, has been rigged in one's favor). The point is that, in addition, an event whose occurrence is, say, *almost* locally necessary (i.e., an event that would occur in *most* close possible worlds but not in *all* close possible worlds) might also count as fortunate if the agent lacks control over it.

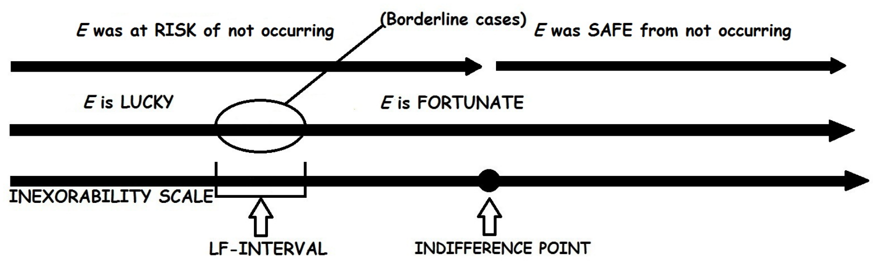
To conclude, the distinction between luck and fortune does not amount to the distinction between risk and safety. Assuming that the lack of control condition does not hold, the difference between lucky and fortunate events is determined by the lucky point in the inexorability scale. By contrast, the difference between risky and safe events is determined by the indifference point. The two points need not coincide.

1.7.2.3 *Borderline Cases*

While there are clear-cut cases of luck (e.g., winning a fair lottery) and clear-cut cases of fortune (e.g., winning a lottery that, unbeknownst to one, has been rigged in one's favor), there are some cases in which we do not know whether to ascribe luck or fortune. Contrary to what one might think, this is no objection to the present account, since as I noted in the preliminary remarks of part I, the ordinary concept of luck is inherently vague and consequently we should not expect of our analysis of luck to remove all vagueness. On the contrary, it would be positive if it could predict the existence of such cases.

Recall the case by Pritchard (2005: 143) that I used to exemplify the point that luck is a vague concept: suppose that *S* drops her wallet, keeps walking and after five minutes returns and finds the wallet in the place where she dropped it. It is by luck that *S* has found her wallet? The answer is not clear.

In the same way, the distinction between luck and fortune needs not be a sharp distinction. We will find borderline cases of events over which we lack control but which are neither clearly lucky, nor clearly fortunate. These cases are located around the lucky point. In fact, it would be more accurate that we conceived the lucky point, rather than as a point, as an interval. We can call it the lucky-fortunate interval (LF-interval for short) and we can represent the idea as follows:



How can we know whether a case is located in the LF-interval? In general, if a case is in the LF-interval, we will be unable to appeal to the modal fragility or robustness of the event in question to explain that is lucky or fortunate. Rather, the only thing to which we can appeal to explain the 'luckiness' or 'fortunateness' of the event is the fact that one lacks control over the event. That is, the somewhat vague limit in the inexorability scale between the technical notions of luck and fortune (the LF-interval) emerges when the lack of control intuition does by itself all the explanatory work.

As we will see in chapter 4, the technical distinction between luck and fortune will do some important philosophical work. In particular, it will be extremely useful for understanding why a true belief is not knowledge. Epistemologists are well aware that the kind of knowledge-undermining phenomenon underlying Gettier-style cases is epistemic luck. My contention is that a heterogeneous group of epistemic cases (some of them Gettier-style), instantiates a different knowledge-undermining phenomenon: epistemic fortune.

1.7.2.4 *Luck and Fortune Defined*

Let us give a couple of working definitions of luck and fortune:

- Luck (luck/fortune-i): *E* is lucky for *S* if only if (1) *E* is significant for *S*; (2) *E* occurs in the actual world but would *not* occur in a large enough proportion of close possible worlds where the relevant initial conditions for *E* are the same as in the actual world; (3) *S* lacks control over *E*.

- Fortune (luck/fortune-ii): *E* is fortunate for *S* if only if (1) *E* is significant for *S*; (2) *E* occurs in the actual world and would occur in a large enough proportion of close possible worlds where the relevant initial conditions for *E* are the same as in the actual world; (3) *S* lacks control over *E*.

These are obviously not definitions of luck and fortune as ordinarily understood. As I have explained, in ordinary discourse there does not seem to be any difference between the two. Instead, the definitions can be seen as Carnapian explications of the concept to which the terms 'luck' and 'fortune' refer when ordinarily used. According to Carnap (1950: 7-8), the task of an explication consists in transforming a more or less inexact concept (the *explicandum*) into an exact one (the *explicatum*) or, alternatively, in replacing the former by the latter. Carnap thinks that there are several requirements that a concept must fulfill in order to be an adequate explicatum of a given explicandum: (1) *similarity* to the explicandum, so that the explicatum can be used when the explicandum is used; (2) *exactness* in the rules of its use (e.g., in the form of a definition); (3) *fruitfulness*, in the sense that the explicatum can be used for the formulation of many universal statements; (4) *simplicity*: the explicatum should be as simple as possible (as simple as (1), (2) and (3) permit).

Let us see whether the above definitions fulfill Carnap's requirements. (1) They are similar to the concept to which the terms 'luck' and 'fortune' refer when they are ordinarily used and can be used in most cases where that concept is used. (2) Besides, the definitions provide quite exact rules of use, particularly when it to comes to the paradigmatic cases: roughly, if the relevant event over which an agent lacks control is modally fragile, the explicatum that applies is luck, whereas if the relevant event is modally robust, the explicatum that applies is fortune. It must be admitted that these two explicata do not remove all vagueness or inexactness from the ordinary concept. As we have seen, they allow for borderline cases. In addition, their reliance on the notion of closeness between possible worlds, which is itself quite vague, introduces some inexactness in their application.

Nevertheless, we should not see this as a defect. As I have argued, the concept to which the terms 'luck' and 'fortune' refer is inherently vague, so that it is positive that the notions used to explain that concept also take into account some of its vagueness. (3) On the other hand, the explicata luck and fortune are very fruitful: they will help us to state interesting generalizations, for example, about the concept of knowledge. (4) Finally, the explicata are quite simple: luck and fortune have been conceived as forms of risk.

§ Some cases reconsidered. We can now explain why cases like Lackey's Buried treasure, Rescher's Big Check and Latus's Rich Uncle have the same structure. On the one hand, the events in question occur in the actual world and would still occur in all close possible worlds, which means that they are safe from not occurring. On the other hand, the agents involved lack control over the relevant events, which means that they are at risk with respect to them. In other words, cases of this sort are cases of fortune, as we have defined it above.

As regards Buried Treasure, I do not deny that Lackey is right in saying that it involves luck in the ordinary sense, since fortune, as I have defined it, is a technical notion that captures one side of the concept to which the terms 'luck' and 'fortune' refer in ordinary discourse. Thus, we can say that Vincent is lucky or fortunate to discover the treasure using the ordinary sense of the terms 'lucky' and 'fortunate' while at the same time we can defend that the phenomenon instantiated by the case is fortune, in the technical sense of term.

The reason Vincent is fortunate but not lucky (in the technical sense of the terms) is not the extraordinary coincidence that the case presents, as I have explained, but the modal robustness of Vincent's discovery and the lack of control that he has over it. Vincent is fortunate in the technical sense because there is no risk that he fails to find the treasure and he is at risk of finding the treasure.[45] If he had known that there was a treasure in the area, we could have brought a metal detector. In that case, we would not have said that Vincent's discovery was by luck or by fortune, because he would have had control over the discovery.

1.8 summary

In this chapter, I have thoroughly analyzed the different ways of conceiving the chance condition for luck and, in turn, the very nature of luck. Let us briefly recap the main points. First, I have emphasized the divergence between the notion of luck and the notion of accident. The main difference between them is that if an action that an agent performs with the intention of bringing about some result is causally relevant for the occurrence of that result, then the occurrence of that result is not an accident although it might be by luck. Second, I have argued that although all coincidences involve luck, some lucky events do not involve any coincidence and, third, that luck is not a matter of indeterminacy: many would accept that there is luck in deterministic worlds (this point might be found controversial by some philosophers, though) Fourth, I have also explained that luck is not a perspectival matter, in the sense that whether an event counts as lucky or not does not depend on our degree of belief concerning its future occurrence or on whether we consider it risky, accidental or indeterminate. Finally, I have argued that neither luck can be defined in terms of epistemic probabilities nor it is an epistemic notion, in the sense that whether an event counts as lucky or not does not depend on whether one is in a bad epistemic position with respect to the future occurrence of that event. One can be in an excellent position to know that an event will occur, deliberately refuse to know it, and the subsequent occurrence of the event still be by luck.

In my view, luck is a special kind of risk. In particular, it involves two kinds of risks: 1) the risk that the lucky event had of not occurring (event-focused sense of risk) and 2) the risk of the lucky agent with respect to the event (agent-focused sense of risk). I have distinguished

45 Remember that the notion of agential risk that I am using can be applied both to positive and negative events. To find the explanation more palatable, imagine that Vincent finds a dangerous mine instead of a treasure.

two legitimate objective ways of conceiving the event-focused sense of risk: a probabilistic and a modal way. I have argued that only the modal one captures the nature of luck (lucky events are not events whose occurrence was not probable, because we have seen that there are highly probable lucky events).

I have explained that events at no risk of not occurring can be thought as inexorable in a given context, where safety is conceived as a point in an inexorability scale beyond which things could not have been otherwise. Below that point, events are to different degrees at risk of not occurring. Importantly, not every actual event that was at risk of not occurring and over which the agent lacks control is a lucky event. Out of control significant events whose occurrence is inexorable or almost inexorable (i.e., not completely safe from not occurring) in a given context involve some 'luckiness' or 'fortunateness'. I called this sort of events fortunate events. Both luck and fortune involve agential risk, i.e., lack of control. The difference between them is that lucky events were at serious risk of not occurring whereas fortunate events were at no or at little risk of not occurring (recall that the distinction is technical). Finally, the (inexact) threshold beyond which an event is no longer lucky and starts being fortunate varies as a function of the significance that the event has for the agent. The question is: how should we understand the relevant notion of significance? This is the topic of the next chapter.

2

THE SIGNIFICANCE CONDITION

2.1 AGENTS WITH INTERESTS

Stones cannot defend themselves from the effects of erosion. However, they are not unlucky or unfortunate for that because they have no interests at all. By contrast, we are certainly unfortunate for suffering the erosion of our bodies through time, because the aging process is something that we cannot control but, more importantly, because it affects our most primary interest: being alive. If some events are good or bad luck or fortune for us, it is because we are beings with interests positively or negatively affected by the eventualities that we cannot control.

It is clear then that not every event that is out of our control is lucky or fortunate: only those with a minimal impact on our interests, i.e. those that are minimally significant for us. For example, the collision of two atoms in a remote area of the universe does not constitute a stroke of luck or of fortune because it has no influence or impact on us whatsoever, not even causal. By contrast, winning a lottery is, for the good or for the bad, something that has a strong impact on our lives.

We are not the only beings that can be lucky and fortunate. Apes, dogs, rodents and even insects are also subject to luck and fortune. Our intuitions are not so strong when it comes to more basic forms of life. For example, can archaea (single-celled microorganisms) be lucky or fortunate? The answer is far from clear. In general, the less complex an organism is, the less intuitive is to ascribe luck to it. The reason is that the less complex an organism is, the less interests it has.

How did interests arise in the first place from non-living matter? Dennett (1984) following Richard Dawkins's best-seller *The Selfish Gene* (1976) explains that interests arose with the emergence of replicators, which were molecules with the power of creating copies of themselves. As replicators started to create an increasing number of different and more complex copies, they developed stable tendencies to preserve certain parameters of their organisms (varieties of homeostasis). Some of those tendencies contributed to preserving and further replicating those organisms within changing environments thus sharpening the definition of their interests in self-preserving and self-replicating. As more and more varieties of homeostasis appeared and prevailed, the range of interests expanded:

> Thus if body-temperature maintenance played an important role in the self-preservation of members of a species,

> body-temperature maintaining control systems that evolved would persist. And the species' catalog of interests would come to include the maintenance of certain (range of) body temperature (...) Food seeking, predator avoiding, mate locating, mating, and health maintaining (self-repairing, trauma avoiding, energy conserving, and so on) are the highest-level subgoals of replicators. In interaction with the particular species' circumstances, these subgoals breed other, instrumental subgoals: odor detecting, hole digging, locomoting, pattern recognizing, pain feeling, mate impressing, and so forth. (Dennett 1984: 22)

§ *Objective vs. subjective interests*. It is needless to tell the story of how the development of complex forms of social interaction and the emergence of culture gave rise to the sophisticated array of human interests. For what matters here, the evolutionary story above tells us that there are interests that humans share with the rest of species: biological *imperatives* or *needs*.

On might think that ascribing interests to very basic forms of life is incorrect, as one might argue that the notion of interest is a cognitive notion and hence presupposes the possession of at least a basic form of psychology. Yet, the ascription of interests to basic forms of life is widespread in biology. By 'interests' here is meant simply the goals of living things. This broad not necessarily psychological sense of the term allows us to say that humans and bacteria share some objective interests, e.g., self-preservation.

Nathan Ballantyne (2012) distinguishes between *objective* and *subjective* interests. For Ballantyne, objective interests encompass interests that have to do with health or with proper biological functioning, but also encompass the kind of interests that are generally relevant to lead a life, like knowledge or friendship (Ballantyne 2012: 331). He thinks that the latter are not subjective because they do not merely consist in desire-satisfaction or pleasurable experience, which are the marks of subjective interests. According to Ballantyne, subjective interests include the objects of mental states like desires and preferences, and also self-consciously adopted goals.

The question is: do lucky and fortunate events affect our subjective or our objective interests? Ballantyne's answer is clear: they have an impact on both of them. Furthermore, one and the same event can have a positive impact on one type of interest while having a negative impact on the other type. By way of illustration, consider the following case:

> Deathly Oscar
>
> A famous director offers Christian the protagonist role in his next film, the shooting of which is planned in a year's

> time. The role that he is supposed to play is that of a skinny machinist who has suffered chronic insomnia for a year. He says yes without thinking it twice, as he knows that roles involving physical transformation increase the chances of winning an Oscar. Christian, who by the time of the offer is shooting an experimental documentary in which he must have three junk food meals per day for thirty days, plans a strict diet plan for the next year so he can lose weight in a healthy way. Just when the documentary finishes, the director of the film desperately calls Christian saying that due to financial pressures they must begin the shooting the next month, so if he manages to look emaciated in a month, he will get the role; if not, the producers will hire a less talented but skinny actor. Christian works out and barely eats for one month but he just returns to his regular weight before the documentary. Out of desperation, he goes to a bar to get drunk. While drinking, he eats some peanuts in a bowl on the bar. One of them happens to be an untested weight-loss pill that a researcher of a pharmaceutical company has accidentally dropped there. Consequently, Christian loses a brutal amount of weight within hours. The next day, the producers do not hesitate in hiring him for the film. In the end, Christian wins an Oscar. The day after the ceremony, he dies because of the secondary effects of the untested weight-loss pill.

The moral of the case is that it is both good and bad luck for Christian to have ingested the pill. It is good luck because his desire (subjective interest) of playing the role and thus of having the chance to win an Oscar is satisfied. Objectively, however, it is bad luck to have ingested the pill because its secondary effects kill him. The case shows that one and the same event might have different impact on one's subjective and objective interests.[1]

§ *The significance condition for luck*. In the literature, there are three main versions of the significance condition for luck.[2] Let us begin with Pritchard's significance condition (2005: 132-133):

- Significance 1: E is lucky for S only if S would ascribe significance to E, were S to be availed of the relevant facts.

1 Obviously, subjects need not be aware of the impact that the relevant events have on their interests in order for those events to be lucky or fortunate for the subjects. Latus (2003: 470, fn. 27) gives a good example: "Someone who is miserable because of the ending of a destructive romantic relationship may well be described as not knowing how lucky he is".

2 The same conditions are also necessary for fortune, as I have defined it.

The main problem that Coffman (2007) and Ballantyne (2012) see with SIGNIFICANCE 1 is that it entails that lucky agents must be capable of ascribing significance, which is problematic insofar as the condition does not allow sentient nonhuman beings (Coffman 2007: 387) and human beings with diminished capacities like newborns or comatose adults (Ballantyne 2012: 324) to be lucky. Coffman proposes an alternative significance condition:

- SIGNIFICANCE 2: *E* is lucky for *S* only if (i) *S* is sentient and (ii) *E* has some objective evaluative status for *S* (i.e., *E* has some objectively good or bad, positive or negative effect on *S*).

Ballantyne (2012: 321) thinks that the second clause should be read as follows:

- (ii)* *E* has some objectively positive or negative effect on the mental states of *S*.

Against this move, Coffman could argue that (ii) does not only entail that lucky events have effect on the individual's mental states. However, Ballantyne anticipates that reply and argues that if the effect is not on the individual's mental states, it is not obvious why (i) is required.

Ballantyne subsequently proposes a counterexample to SIGNIFICANCE 2 in which an unlucky man has no inkling that scientists have randomly selected him to put his brain in a vat that feeds his neural connections with real-world experiences. The case is allegedly problematic for SIGNIFICANCE 2 because the event, which is bad luck for the man, has no impact on the man's mental states and, in particular, on his interior life, which is not altered.

In reply, Coffman could argue the following: although the fact that the man's brain is being put in a vat does not have effects on the man's interior life and namely on his phenomenal mental states, it certainly affects his representational mental states, as most of them turn out false. Accordingly, SIGNIFICANCE 2 can predict why the man is unlucky. For this reason, I do not think that Ballantyne makes a case against SIGNIFICANCE 2. Nevertheless, I do think that Ballantyne's significance condition is superior to Coffman's:

- SIGNIFICANCE 3: *E* is lucky for *S* only if (i) *S* has an interest *N* and (ii) *E* has some objectively positive or negative effect on *N*.

SIGNIFICANCE 2, as it stands, does not tell us the sort of attributes of individuals that are supposed to be affected by lucky and fortune events. Do lucky events affect the qualitative states of sentient beings? Do they affect their representational states too? Is the physical condition of individuals also affected? SIGNIFICANCE 3, by contrast, is more specific in this regard: what lucky and fortunate events affect are the subjective and objective interests of individuals.

2.2 LUCK AND FORTUNE REDEFINED

We can implement the working definitions of luck and fortune given in chapter 1 with SIGNIFICANCE 3:

LUCK: *E* is lucky for *S* if only if:

(1) *S* has an interest *N* and *E* has some objectively positive or negative effect on *N*,

(2) *E* occurs in the actual world but would not occur in a large enough proportion of close possible worlds where the relevant initial conditions for *E* are the same as in the actual world,

(3) *S* lacks control over *E*.

FORTUNE: *E* is fortunate for *S* if only if:

(1) *S* has an interest *N* and *E* has some objectively positive or negative effect on *N*,

(2) *E* occurs in the actual world and would occur in a large enough proportion of close possible worlds where the relevant initial conditions for *E* are the same as in the actual world,

(3) *S* lacks control over *E*.

I argued that luck and fortune are special forms of risk. LUCK shows that luck is the combination of two forms of risk: the risk that the lucky event had of not occurring (2) and the risk that the lucky agent has with respect to the event (3), which is significant for her (1). On the other hand, FORTUNE shows that fortune basically arises from one type of risk: the risk that the fortunate agent has with respect to the event (3), which was at no or at little risk of not occurring (2) and which is significant for her (1).

2.3 AGENTS AT RISK

As we have defined the agent-focused sense of risk, an agent is at risk with respect to a significant event just in case the agent lacks control over the event. In this chapter, we have learned that significant events are events that have an effect on our subjective and objective interests, an effect that may be positive or negative (this determines whether those events represent strokes of good or bad luck/fortune for us).

Do an agent's interests also play a determinant role in her being at risk or not with respect to an event? What is the relation between risk and interests? We know now that the significance that an event has for us is understood in terms of the (positive or negative) impact that the event has on our interests. Accordingly, not every event that is beyond our control will be risky for us precisely because not every event is significant for us. For example, for most of us, the fall of a leaf in the middle of the Amazonas has no impact on our objective and subjective interests, as it neither affects our health, nor our biological functioning, nor has an effect on the objects of our desires and of our preferences. In the same way, we are at no risk with respect to several atomic nuclei joining in a very distant point of the universe (something that we cannot control). However, we are certainly at risk with respect to the same nuclear fusion if it triggers off a nuclear explosion near us (something that we cannot control either). The reason is clear. In the latter case, the nuclear fusion affects our most important objective interest: being alive.

At this point, we can redefine AGENTIAL RISK, the technical notion of risk that I have used to define luck and fortune (the one that condition (3) captures), by making explicit the connection between risk and interests:

- AGENTIAL RISK: *S* is at risk with respect to an event *E* if and only if (i) *S* has an interest *N*, (ii) if *E* were to occur, it would have some objectively positive or negative effect on *N* and (iii) *S* lacks control over *E.*

As I noted in section 1.6.4, AGENTIAL RISK diverges from the ordinary notion of being at risk with respect to an event. In ordinary discourse, the term 'risky' is used as synonymous of 'dangerous', where a dangerous event is an event with adverse or unwelcome consequences.[3] We can give the following definition of danger and hence of risk as ordinarily understood:

- DANGER: *S* is in danger with respect to *E* if and only if (i) *S* has an interest *N*, (ii) if *E* were to occur, it would have some objectively negative effect on *N* and (iii) *S* lacks control over *E*.

If people do not ordinarily say that lottery players are at risk of winning the lottery, it is because they have a conception of risk as danger in mind. AGENTIAL RISK allows for a technical usage of the expression 'at risk of', which not always means 'in danger of'. In general, AGENTIAL RISK says that we are at risk with respect to whatever significant

3 In section 1.6.4, I explored the connection between significance and risk. The two main ideas exposed there about how significance is related to risk can be stated now by making explicit the idea that the significance of an event for an agent is a matter of how much the event affects the agent's interests. The two main ideas of that section can be thus restated as follows: 1) the impact that an event has on an agent's interests is one of the pragmatic factors that help to determine the position of the indifference point on the inexorability scale (the point beyond which an event is no longer at risk of occurring —or of not occurring); 2) the greater the impact that an event has on an agent's interests, the smaller needs to be the proportion of close possible worlds in which the event would occur to be considered *dangerous* for the agent.

event is beyond our control, which entails that we can be at risk with respect to events that increase our well-being.

In which sense are these events risky for us? We are at risk with respect to them in the sense that they affect us in ways that diverge from the path traced by our goal-directed controlled actions. When we have control over an event, we can count on it to perform other actions. However, events beyond our control, even those that carry positive effects, are events on which we cannot count to take further action.

For instance, inexperienced investors playing the stock market typically buy shares without knowing the relevant financial technicalities or the maneuvers that big investor groups perform to make money at their expense. Even if the prices of the shares rise and inexperienced investors become rich (which is positive as far as their subjective interests are concerned), the rise of the prices is something on which they cannot count at the time of the investment because it is something beyond their control.[4] Thus, it would be irrational for them to apply for a big loan from a bank to set up an expensive business on the assumption that the forthcoming profit in the stock market will be sufficient to repay it. AGENTIAL RISK allows for good risks, but risks after all and, in general, we cannot rely on what is risky for us, even if unknowingly beneficial.

2.4 SUMMARY

This chapter has been devoted to the significance condition for luck and fortune. In section 2.1, I have analyzed several significance conditions. I have argued that the best condition in the literature is Ballantyne's, which defines significance in terms of the impact of the lucky event on the agent's interests, where the kind of interests at issue are subjective (e.g., desires and goals) and objective (e.g., biological functioning). In this way, events are lucky or fortunate for us because they affect our interests. In section 2.2, I have implemented the working definitions of luck and fortune given in section 1.7.2. In section 2.3, I have further explained that lucky and fortunate events not only affect our interests, but they affect them in a risky way. Even if the consequences of lucky and fortunate events are positive, they are still risky for us because their impact is something that we cannot control. In this sense, we are only safe with respect to which we can control. The question is: how should we understand the relevant notion of control? This is the topic of the next chapter.

4 I will specify in the next chapter in which sense of the term 'control' inexperienced investors lack control over their investments.

3 THE LACK OF CONTROL CONDITION

Let us begin with a brief review of the literature. As Pritchard (2005: 127) points out, the most widespread idea concerning luck that we can find in the literature is perhaps the idea that events occurring as a matter of luck are events beyond our control:

- LACK OF CONTROL: *E* is lucky for *S* only if *S* lacks control over *E*.

The origins of the idea can be traced back to Thomas Nagel's famous characterization of the phenomenon of moral luck:

> Where a significant aspect of what someone does depend on factors beyond his control, yet we continue to treat him in that respect as an object of moral judgment, it can be called moral luck. (Nagel: 1979: 26)

In the huge literature on moral luck, the very concept of luck is commonly accounted for in terms of lack of control (see e.g., Greco 1995; Moore 1990; Statman 1991 and Zimmerman 1987). In the less extensive literature on the general notion of luck, a representative group of commentators think that a lack of control condition is necessary for luck (Coffman 2007, 2009; Latus 2003; Levy 2009, 2011; Riggs 2007, 2009), with the significant exception of Pritchard (2005), who affirms that the condition is not required. In the specific literature on epistemic luck, Pritchard thinks, accordingly, that a lack of control condition is not needed to define epistemic luck, while Riggs (2007, 2009) explicitly holds the opposite. Statman (1991) also suggests that epistemic luck, as well as moral luck, arises from factors beyond the agent's control.

§ Methodological issues. Ascriptions of control are made about all sorts of things: we say that we control cars, our emotions, animals, the volume, passports or the crime rate. What we mean with such ascriptions is plausibly that we control behaviors, events or states related to those things. In philosophy, the term 'control' has been used extensively in accounts of a variety of concepts: action, property and ownership, freedom, privacy, personal autonomy, responsibility, luck.

In the light of the many ordinary and philosophical applications of the term 'control', it is not surprising that LACK OF CONTROL, as it stands, is silent on the conditions under which someone counts as having control over something or someone. This is not a defect, though. For LACK OF CONTROL is part of a general definition of luck that must be generic enough to account not only for the ordinary notion of luck but also for all the varieties of luck that philosophers distinguish in their arguments (e.g., distributive luck, moral luck, epistemic luck),[1] and also for their sub-varieties (e.g., constitutive luck, resultant luck, veritic luck). For example, the kind of control that is required to achieve moral responsibility is surely different from the kind of control required to achieve knowledge. In a definition of epistemic luck, we would not accept a notion of control that requires direct voluntariness of belief (i.e., that entails that beliefs can be formed upon a direct act of the will), whereas in a definition of moral luck we would accept with a conception of control that involves voluntariness of action. In sum, the best way to grant without much controversy the universality of our definitions of luck and fortune is to leave LACK OF CONTROL unspecified.

Of course, this does not mean that we cannot specify the notion of control to account for the different varieties of luck and fortune that are interesting for philosophers. But we cannot assume without argument that all species of luck, whether moral, ordinary, epistemic or distributive, can be properly defined by understanding the relevant lack of control condition in the very same specific manner.

In section 3.1, I will illustrate why it is an error to proceed under the assumption that an specific definition of control will apply to any variety of luck. In particular, I will show why E. J. Coffman's specification of LACK OF CONTROL in terms of not being both free to produce an event and free to prevent it cannot be part of a generic definition of luck (it makes the definition inapplicable to epistemic luck). In addition, I will also argue that Coffman's lack of control condition is not necessary for luck. In section 3.2, I will give a general account of the notion of control, which will serve, in section 3.3, to explain a series of cases that Jennifer Lackey (2008) has proposed as counterexamples to the view that luck is essentially a matter of lacking control over an event.

3.1 IS CONTROL A MATTER OF CHOICE?

According to Coffman (2009), an event *E* is significantly beyond *S*'s control if and only if *S* is not both free to do something that would (non-redundantly) help produce *E* and free to do something that

1 Distributive luck is the kind of good or bad luck that arises when one has more or less opportunities, resources or welfare simply in virtue of, for instance, one's birthplace, (think about children born in dysfunctional families). Moral luck arises when luck (e.g., luck in the consequences of one's actions) makes a moral difference (e.g., a difference in moral responsibility or in moral justification). Finally, epistemic luck is the kind luck that concerns epistemic factors such as how an agent comes to believe the truth.

would (non-redundantly) help prevent *E*.[2] Accordingly, he specifies LACK OF CONTROL in the following way:

- LACK OF CHOICE: *E* is lucky for *S* only if *S* is not *both* free to do something that would help produce *E and* free to do something that would help prevent *E*.

To understand better LACK OF CHOICE, let us consider several meanings of the expression 'being free to' or of its equivalent expression 'having a choice about'.

§ Senses of 'having a choice about'. Erik Carlson (2000) distinguishes several senses of being free to do φ or of having a choice about φ or, as he puts it, of having a choice about a truth *p*. I will consider only four of the eight senses that he distinguishes, as the other senses are formulated in terms of might-clauses rather than in terms of would-clauses, which are the ones applicable to LACK OF CHOICE. The four different senses are:

S has a choice about a truth *p* if and only if:

1. *S* is able to act so that *p* would be false.
2. *S* is able to act in a way that would causally contribute to *p*'s being false.
3. There is way of acting, such that (i) *S* is able so to act and (ii) *S* knows that if he were so to act, *p* would be false.
4. There is way of acting, such that (i) *S* is able so to act and (ii) *S* knows that so acting would causally contribute to *p*'s being false.

1 and 2 are non-epistemic senses, whereas 3 and 4 are (partially) epistemic. To illustrate the distinction, Carlson (2000: 281) gives the following example. Suppose that *S* does not know the combination that opens a locked safe. Carlson asks, does *S* have a choice about whether the safe will remain locked for the next few minutes? In a non-epistemic sense of 'having a choice about', *S* does have a choice because she has working hands and therefore she is able to dial any combination. However, people with a more demanding sense of 'having a choice about' in mind will deny that *S* has a choice about whether the safe will remain locked. For them, *S* has no choice simply

2 With the expression 'non-redundantly', Coffman intends to avoid regarding as cases of control certain cases of causal overdetermination, e.g., a case in which one contributes to an ongoing avalanche by throwing a snowball into it. Although the avalanche is out of one's control, one has the option of making a causal contribution to it. This way of conceiving control is very similar to John Martin Fischer and Mark Ravizza's notion of regulative control, which is the kind of control that requires access to alternative possibilities (freedom to choose and do otherwise). See Fischer & Ravizza's seminal work (1998) or Fischer (2012) for a more recent contribution.

because *S* she does not know the right combination, i.e., *S* does not *know how* to open the safe.

According to Coffman, Lack of Choice should be read in terms of the epistemic sense of 'being free to'. He argues that if the condition were understood in the non-epistemic sense, it could not explain why cases like the following are cases of luck. Suppose that *A* decides to give *B* one million if either *B* raises her hand within the next five seconds or her next coin flip lands heads. *B* knows nothing about *A*'s plans. Suppose, in addition, that *B* does not raise her hand and that the coin lands heads. Intuitively, *B* is lucky to win the money. The problem is that if we understand 'being free to' along the lines of 1 or 2, we are committed to say that *B* is free to or has a choice about winning the million, as she is able to raise her hand. In view of that, Coffman opts for the epistemic reading: since *B* does not know that by raising her hand she will win a million, she has no choice about whether she will win.

§ Objection 1: the condition is inapplicable to epistemic luck. Lack of Choice is not suitable to account for epistemic luck. It would follow from an epistemic version of Lack of Choice that if an agent is free to do something that would help produce, say, that she comes to believe some truth and free to do something that would help prevent that, then her belief is not luckily true and thus it could qualify for knowledge. However, it does not seem that the antecedent of this conditional is applicable to most of our beliefs, as most of the times we have no other option but to form them. This is evident in the case of perceptual beliefs.

Suppose that your hands are before your eyes and that you are staring at them. Your visual system is in good shape and nothing blocks your line of sight. Are you free to do something that would make you believe the truth? Do you have the choice of not believing that you have hands? Are you free to refrain from believing that? You cannot do anything but to form the belief that you have hands because the experience as of hands makes your belief, so to say, irresistible. This is an essential feature of our senses: they are such that we accept their deliverances at face value when the conditions are normal.

One could reply that, by indirect means, one has the option of 1) not believing that one has hands at will (e.g., by turning off the lights or by cutting off one's hands) and of 2) making something that would help produce one's coming to believe truly that one has hands (e.g., by cutting off one's hands and by re-implanting them). One might have such options, true. However, would it be correct to claim that the lucky or non-lucky status of a belief depends on such actions? The answer is far from affirmative.

§ Objection 2: counterexample. Fair Lottery Player, the counterexample I proposed to Steglich-Petersen's epistemic chance condition is

problematic for Coffman's lack of control condition too. Recall the case:

> FAIR LOTTERY PLAYER
>
> John is having serious financial problems. Desperate as he is, he thinks it would be a good idea to buy a Megalotto ticket, which on that occasion has a 50-million jackpot. The peculiarity of Megalotto is that a random system extracts the winner ticket before the tickets are sold. After that, lottery workers put the tickets into envelopes that are randomly distributed.
>
> John goes to buy a lottery envelope to the grocery store of his friend Jim. Jim is aware of John's financial problems, so he tells John: "Look, buy this envelope: the winner ticket is inside". John asks Jim how can he know that, to which Jim answers that he knows it because his cousin, a lottery worker, has told him that he has put the winner ticket inside that specific envelope.
>
> However, Jim's cousin was just playing a joke to Jim. The strict lottery system prevents workers from knowing where the winner ticket is. Suddenly, Jim receives an emergency call and must run to the hospital immediately. He asks John to stay in charge of the store while he is absent. And there he is John, before the envelope, having the chance of checking whether the winner ticket is inside. But John deliberately refuses to open the envelope. "Honesty above all", he thinks. When Jim returns, John buys the envelope. Luckily, the winner ticket is inside. John wins the lottery.

The case poses a problem to LACK OF CHOICE because, on the one hand, John is free to make something that would prevent him from winning the lottery: not buying an envelope. On the other hand, John is free to perform actions that would bring about his winning. In particular, he is free to open the envelope. By opening the envelope, he would know that the ticket is inside, which means, in turn, that by subsequently closing the envelope and buying it he would win the lottery. The problem is that John refrains from performing such actions and takes a third course of action: he buys the envelope without knowing what ticket is inside. On the other hand, since the winner ticket is inside, he wins by luck. Therefore, John is lucky to win even if he is both free to do something that would make him win and free to do something that would prevent it. Consequently, LACK OF CHOICE is not necessary for luck.

However, Coffman considers a similar case and anticipates the objection. Coffman's point is that the agent, in our case John, is free to perform an action that would result in an illegitimate win but he is not free to perform an action that would result in a *legitimate* win. Accordingly, winning the lottery *fairly* is a stroke of good luck for John because he has no choice about it and thus Lack of Choice would hold.

In reply, I do not see why we should disambiguate the expression 'winning the lottery' in the way Coffman suggests. When we ascribe luck to a lottery winner we do not necessarily have in mind that he is lucky to have won fairly. We might have in mind that he is lucky to have won a big amount of money, *simpliciter.* At any rate, assuming that such a disambiguation is in place, we can still block Coffman's response by tweaking Fair Lottery Player:

> Fair Lottery Player II
>
> The lottery system gives people with financial problems the following option. Before the lottery draw takes place, the organizers send to them a very hard Sudoku puzzle, whose solution is the number of the winner ticket. The first person to solve the Sudoku is called by the organizers, who tell her where the envelope with the ticket would be sent. In addition, there is another prize (with the same amount of money) so that players without financial problems can have their chances of winning too. John, who has financial problems, cleverly solves the Sudoku, so that he comes to know the winning number and the location of the winner ticket. He goes to the shop where the envelope with the winner ticket is (all the envelopes have the number of the ticket inside written on them). When he is about to buy the envelope, a very poor man enters in the shop. In an act of extreme generosity, John tells the man to buy the envelope where he knows the winner ticket is. The poor man wins. After that, John buys another envelope. "Who knows", he thinks, "perhaps I win the other prize". He opens the envelope and, luckily, the second winner ticket is inside.

John is free to perform an action that he knows would result in a *legitimate* win. However, he refuses to perform it and he wins the lottery by luck *in a fair way*. Coffman could reply that John is free to win the first but not the second prize. To block that answer, we could further complicate the case by making John win the first prize by luck: a secret rule of the lottery says that if the person who solves the Sudoku (John) cedes the winner ticket to a stranger, the stranger will receive certain amount of money and a coin will be tossed to decide whether that person keeps the first prize or not. John, luckily, wins the first prize in this way.

In sum, Fair Lottery Player II shows that an event might be by luck even if one is free to act in a way that one knows would bring about (or would prevent) the occurrence of the event. The moral of the case is that in order for one to prevent an event from being by luck one must take effective action. In addition, the moral of this section is that if we want to account for the general notions of luck and fortune, the best thing we can do is to leave Lack of Control unspecified. Of course, this does not mean that we cannot give a general account of the notion of control or that we cannot formulate versions of Lack of Control to account for specific forms of luck and fortune. But philosophical prudence compel us to make as few assumptions as possible in our generic definitions of luck and fortune.

3.2 A GENERAL ACCOUNT OF CONTROL

Until recently, discussions about moral and epistemic luck have used the concept of luck as an unexplained explainer. Something similar occurs with the concept of control. Although in many philosophical arguments the concept plays a crucial role, not much is said about it. Typically, it is assumed that we all can distinguish in an intuitive way when something is under or beyond our control, in such a way that control is regarded as a primitive notion to which one can resort to explain other concepts. Nevertheless, definitions and explanations of the concept of control are sometimes given. Yet, they just aim to clarify its concrete role in the wider philosophical argument where the concept is used.[3] Of course, there is nothing wrong with using the notion of control in some specific sense to serve some philosophical purpose, e.g., to define a specific variety of luck. Indeed, to explain the nature of knowledge, a specific sense of control needs to be distinguished. However, for the sake of clarity, it is useful to start with a proper account of the notion of control in general, as we ordinarily use it.

3.2.1 *Control, Interests and Goals*

It is not easy to find definitions of control outside the scope of philosophical arguments, but Dennett, in one of his works devoted to the compatibility between determinism and free will, gives one:

3 I have in mind, for instance, Fischer and Ravizza's distinction between regulative and guidance control and the subsequent use of the notion of guidance control to support a semi-incompatibilist position about free will (Fischer & Ravizza 1998; see also Fischer 2012). Another example: in a recent paper, Joseph Raz (2011) aims to show that the common principles of responsibility which delimit responsibility to intentional actions (Intention Principle) and to outcomes under our control (Control Principle) are false. Interestingly, before entering into the details of the argumentation, Raz says that "'Control' is used in a context-sensitive way, and there is no need here to explore the notion, except as it is use in the Control Principle", which roughly means "being moved and guided by reasons as one sees them" (Raz 2011: 81-82).

- *A controls B* if and only if the relation between *A* and *B* is such that *A* can *drive B* into whichever of *B*'s normal range of states *A wants B* to be in. (Dennett 1984: 52)

Let me start with a very basic and intuitive analysis of the main features of Dennett's definition. After that, I will make a useful distinction between two ways in which we ordinarily apply the term 'control'.

§ *Goal-directedness*. Dennett says that control requires something like desires (he puts the word in single quotes). We can generalize. Control requires a sort of intentionality. The notion of intentionality is typically defined as aboutness or directedness. A stone has no intentionality because it is not about or directed towards anything. By contrast, mental states such as beliefs, fears or wishes are intentional because they are about or directed towards objects and states of affairs. When we control something or someone, it is our intention, desire, goal, aim, target, plan or purpose to achieve certain outcome concerning that thing or person, i.e., the practices or activities in which we engage and the actions we take when we exert control are always *directed towards* a goal or an aim, which may be more or less specific depending on the case. Therefore, the sort of intentionality without which control cannot arise is *goal-directedness*.

§ *Controllees*. There is control only if there is something or someone to control, what Dennett calls the *controllee*. The variety of controllees is wide: people, artifacts, actions, animals, beliefs, utterances, biological processes, and so on. Consider the following list of examples: the media control the public opinion, the government controls the imported goods, drivers control their cars, soccer players control the ball, yogis control their breathing, boxers control their punches, dog owners control their dogs, doctors control the health of their patients, politicians control the tone of their speeches, and so on. Almost anything can be subject to control.

§ *Controllers, goals and interests*. Attributions of control do not typically make transparent the goals of the *controller*. By way of illustration, the attribution "Mary's parents control her life" may imply different goals. For example, Mary's parents may be interested in preventing her daughter from going out at night more than once a week, or perhaps in preventing her from hanging out with certain people, or maybe in making sure that she goes well at school. In the same way, when we hear a worker complaining that the boss is constantly controlling him or her at work, we learn nothing, unless more is said, on why the boss is acting like that. Possibly, the boss has the aim of increasing the quality or the efficiency of the worker, or maybe of increasing the monthly profit at all costs; perhaps, the boss has a non-professional purpose.

In most cases, context makes explicit the relevant goals. The moral to be drawn is that it is possible for us to have control insofar as we are beings with interests. Our interests, from the most basic (e.g., self-preservation) to the most sophisticated ones (e.g., aesthetic and philosophical interests), give shape to our goals, and the actions, practices and processes that give rise to control are directed towards those goals. To have goals or aims is at the core of what is to be a controller.[4]

What counts as a controller? Everything with a goal or an aim can be a controller. All forms of life have goals, from humans and the great apes to single-celled organisms, which have developed stable chemical processes that contribute to their replication and preservation, which are the most primary goals of any organism. The more complex the catalog of interests of an organism is, the more complex forms of control will the organism develop.

However, goals and aims are not exclusive of organisms. Artifacts have aims as well. For example, thermostats aim at regulating the temperature of the environment; the autopilot of a plain aims at maintaining certain trajectory; a computer firewall aims at keeping a network secure. It is an open question whether the goals of artifacts are our goals or not, i.e., whether artifacts have original or derived intentionality. For the purposes here, insofar as they have goals it makes sense to say that a thermostat controls the temperature, an autopilot the trajectory of a plane or a firewall the traffic of a network.

§ Goals and sub-goals. Control may involve simple goals, as when one intends to raise one's arm, which may suffice to control the basic action of raising one's arm, but also complex goals, as when a driver aims at driving from one location to another safely. To control the car and drive safely, the driver must control the direction of the car, the speed, the gas level, the performance of other drivers, and so on. In this way, complex goals involve sub-goals, i.e., intermediate or simultaneous steps that jointly contribute to the achievement of a goal. The more complex a goal is, the more difficult is to exert control (because more parameters are to be controlled).

§ The environment has no goals. If control requires the controller to have goals, then the environment cannot control us. Dennett (1984) attributes to B. F. Skinner the view that the environment can make us do what it wants. However, as Dennett correctly points out, that is nonsense, since the environment aims at nothing. Sometimes, we say things like "The Sea wants to destroy the ship" or "The Mountain does not let us climb it". However, these claims are prosopopeial. In the same way, we sometimes find in philosophy expressions like 'friendly circumstances' or 'environment cooperation', which mislead-

4 Consequently, the term 'wants' in Dennett's definition of control should not be read as referring to a desire but, more generally, to a goal.

ingly imply the idea that the environment or the circumstances can have goals and thus be benign or malevolent to us.

§ *Luck and fortune require interests; control requires goals*. Let me make one final remark. In the last chapter, we interpreted the term 'interests' very broadly, as the goals or aims of living things (see section 2.1). In this sense, interests are types of goals. But there might be goals that are not interests (e.g., the goal of a software program). While control requires goals in general, luck and fortune only require interests (i.e., a type of goals). This is the reason why thermostats cannot be subject to luck or fortune: they do not have interests. We can imagine borderline cases of highly sophisticated robots with goals that mimic the interests of humans or other animals. In those cases, we hesitate before rejecting that robots are subject to luck or fortune. However, our hesitation is only a proof of the point: we are not sure whether highly sophisticated robots are subject to luck and fortune because we are not sure whether they count as forms of life. The question of what individuals have interests is irremediably tied to the question of what individuals are alive, a question that goes beyond the scope of this dissertation.

In what follows, I will distinguish between two senses of the notion of control (i.e., I will distinguish two types of control to which the term 'control' indistinctively refers when ordinarily used). The distinction and large part of the discussion is inspired by some ideas that can be found in the extensive literature on control theory, which is a highly multidisciplinary branch of engineering that studies the formal principles, methods and tools for the design and analysis of engineering systems (physical and informational) capable of achieving established goals by detecting and automatically adapting to changes in the environment (Murray 2003).[5] My aim here is to generalize some of those ideas and use them to account for the notion of control from a philosophical point of view.

3.2.2 *Effective Control*

According to Dennett, *A* controls *B* if and only if *A can drive B* into a desired range of states. Two points. First, in order for *A* to have control over *B* it does not suffice the mere capacity or disposition to drive *B* into certain states. The reason is that *A* might have the capacity to drive *B* into whichever state *A* wants and yet refuse to do it (this was the main problem of Coffman's definition of control). In that case, although *A* has the disposition to control *B, A* does not *de facto* control *B*. Second, what does 'drive' mean?, that is, what is the

5 See also Åström and Murray (2008) for an accessible introduction to the field of control in science and engineering.

nature of the control relation? In most cases, the relation is causal: *A* controls *B* by causing *B* to be in certain state. However, some uses of the term 'control' may not allow for causation. Suppose that it is acceptable to say of some mental event that it controls a physical event. Some philosophers might not be willing to qualify the relation between these events as causal (perhaps they would prefer to qualify it as a relation of determination). After these preliminary remarks, I will call the first sense of the notion of control *effective control*:

- Effective Control: *A* has effective control over *B* if and only if (i) it is *A*'s aim that *B* is in certain state *S*, (ii) *A* has a disposition to cause/determine *B* to be in *S*, (iii) *B* is in *S* and (iv) *B* is in *S* because of *A*'s disposition to cause/determine *B* to be in *S*.[6]

A driver safely driving her car has effective control over her car because (i) she has the aim and (ii) the disposition to maintain or modify the trajectory of the car, its speed, and so on. (a disposition that must be stable and integrated with the other driver's dispositions), (iii) those parameters are as the driver wants and (iv) the fact that they are as the driver wants is because of the driver's disposition to bring them to that state.[7]

3.2.3 *Tracking Control*

Consider now the following possible situation. *A* has the relevant aim and the disposition to drive *B* into certain state; *B* is already in the state *A* wants *B* to be; *A* has done nothing to drive *B* into that state (i.e., *B*'s being in that state is not because of *A*'s relevant dispositions). *A* does not clearly have effective control over *B* (condition (iv) is not satisfied). The important question is the following: can *A* still have control over *B* even if *A* does not have effective control over *B*? As the following examples show, the question has a positive answer:

1. Suppose that the NASA sends an astronaut to the moon. The launch, the trajectory, the speed and the landing of the spacecraft have been carefully planned by the engineers so that if all the parameters are as expected, the systems of the spacecraft will lead it to the moon without need that the astronaut intervenes. However, if an unforeseen event changed, say, the trajectory of the spacecraft, the astronaut would correct it. When everything goes as planned, does the astronaut have effective control over the spacecraft's trajectory? It does not seem so, because the astronaut is not exerting any causal influence on it. However, would we say that the astronaut has no control

6 Depending on the case, one may want to substitute the phrasing 'it is *A*'s aim that *B* is in certain state' by 'it is *A*'s aim that *B* occurs', 'it is *A*'s aim that *B* ceases to occur', 'it is *A*'s aim that *B* gains such and such a property', and so on. In addition, if one finds acceptable some way of specifying the control relation different from causation or determination, one can adapt or extend the definition accordingly.

7 The notion of effective control will be further explained in chapters 6 and 7.

over the trajectory whatsoever? No, we would not, and the reason is that if some parameter (trajectory, speed, and so on) changed, the astronaut would detect it and would intervene if needed.

2. The aim of a thermostat is to maintain the temperature of the environment near a desired setpoint. An important part of a thermostat is its thermometer, which is responsible for registering changes in the environment's temperature. Suppose that this morning the thermostat has not activated the heating yet because the environment's temperature has been naturally raised by the heat of the sun to the desired setpoint. The thermostat has no effective control on the environment's temperature because it has no causal influence on it (at least until yet). Would we say that it has no control over the temperature whatsoever? It does not seem so, since the thermostat, by means of its thermometer, would detect any change of temperature and would activate the heating if needed.

3. The main concern of a doctor is to keep her patients healthy. When a patient gets ill, the doctor not only applies the most adequate treatment to the patient, but she also constrains her patient to adopt certain habits that will eventually heal her. When a patient is healthy, the doctor does not apply any treatment or impose any constraint and, in this way, does not exert any causal influence on the parameters that determine the health of her patient. Would we say that the good doctor has no control over her patient's health if she keeps an eye on the patient's lifestyle and eventually runs tests that make her know her patient's current state? It does not seem so, if anything were abnormal, she would give her patient proper medical care.

In the light of these examples, we can distinguish another form of control. What do the doctor, the thermostat and the astronaut have in common? They have in common that they *monitor*, respectively, the patient's health, the environment's temperature and the spacecraft's trajectory and speed. This is what allows them to have control over these parameters even if causal influence over them is not sustained. In general, *A* may still have control over *B*, even if *A* does not have sustained causal influence over *B* as long as *A monitors B.* I call this form of control *tracking control:*

- Tracking Control: *A* has tracking control over *B* if and only if *A* monitors *B.*

§ *Monitoring.* Monitoring has two components: an epistemic and a dispositional component. When *A* monitors *B*, *A* keeps track, registers or gathers information about *B*. This is the *informational component* of monitoring. In addition, the information that *A* registers about *B* serves *A* to initiate, stop or continue some performance or action that contributes to the achievement of some goal; in a sense, the information compiled disposes *A* or puts *A* in a position to do such things.

This is the *dispositional component* of monitoring. When only the first component is in place we may say that *A* carries out a *merely informational monitoring* of *B*. This is the case, for example, of a thermometer, which merely registers the temperature of the environment, or of an eventual eavesdropper, who just wants to find out what other people are saying. Ascriptions of control might be true when *A* carries out a merely informational monitoring of *B*. This would apply to ascriptions such as 'the thermometer controls the temperature' or 'the eavesdropper controls the conversation'.

Nevertheless, for the most part, monitoring is not merely informational but also *dispositional*, as when the thermometer of a thermostat tracks the environment's temperature in such a way that, depending on how near is the temperature value to the desired setpoint, the internal mechanism of the thermostat keeps, stops or initiates the heating, or as when a spy gets some relevant information with the purpose of overthrowing a government.

3.2.4 *Effective and Tracking Control Combined*

In many systems (mechanical, biological, institutional or virtual), the dispositions responsible for effective control are integrated with the dispositions responsible for monitoring (tracking control) in such a way that they jointly contribute to the achievement of goals. In those systems control may follow two kinds of patterns: *feedforward* and *feedback*.[8] Both feedforward and feedback are ways of dealing with disturbances, i.e., with events that have the potentiality of disrupting the attainment of the system's goal. The essential difference between feedforward and feedback is the time when the monitoring takes place, i.e., the time when the relevant information is gathered. In a feedforward system, the information is gathered *before* the activity takes place. In a feedback system, the information is gathered *after* the systems acts.

Feedforward may be hard to implement, because it requires foreseeing disturbances so that the system can avoid them. For this reason, it is suitable for circumstances in which disturbances can be predicted. Feedback, by contrast, can operate in circumstances where the disturbances are unpredictable, because it operates after they take place regulating in this way the performance of the system. In addition, feedback and feedforward are not incompatible but complementary. A system may implement both forms of control to achieve some goal.

To illustrate the difference between both forms of control, consider a common car's cruise control system. A cruise control system has the aim of keeping the speed of the car at a constant value. To achieve that aim, cruise control systems are typically equipped with speed

8 See Åström & Murray (2008: chapter 1) and Hopgood (2012: chapter 14) for relevant discussion and some illustrative examples of feedback and feedforward systems.

sensors installed in the car's wheels, so that if the speed is lower than the desired value, the cruise control accelerates the car; if the speed is higher, it decelerates the car; if it is the desired one, it keeps the *rpm* constant. This way of controlling the car's speed is by feedback.

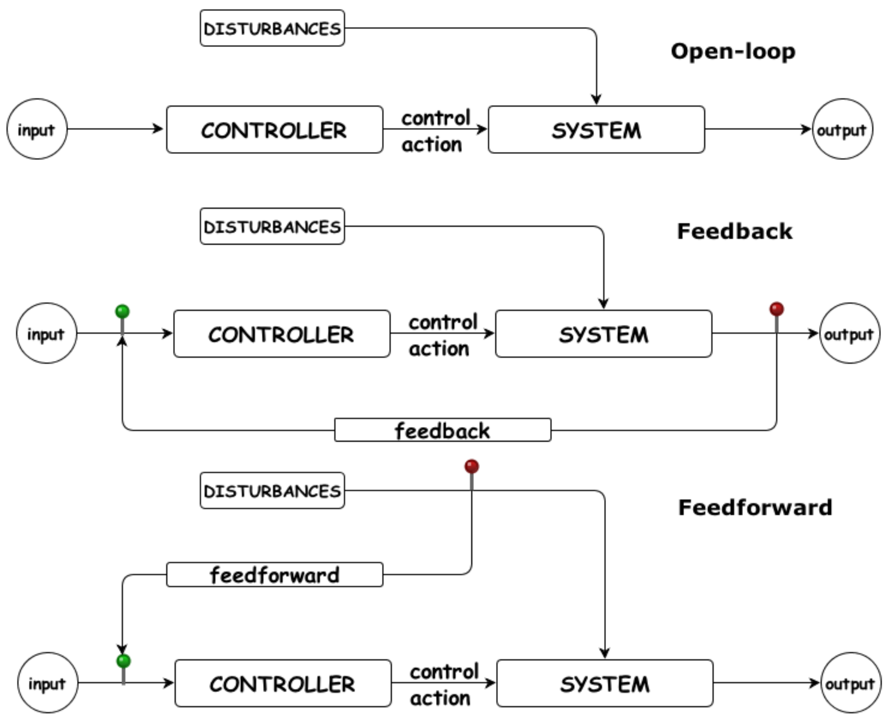
On the other hand, we can imagine a cruise control system equipped with a device capable of detecting the ascents and the descents of the road in advance (e.g., with a GPS). If the device detects a point in the road with a significant change of inclination, the cruise control will increase or decrease the *rpm* of the motor before the car arrives at that point so that the speed can stay stable. This way of controlling the car's speed is by feedforward.

Let us consider again the cases used to exemplify the notion of tracking control. Thermostats have feedback control over the environment's temperature, because they detect any temperature change and increase or decrease the temperature until it reaches the desired value. On the other hand, NASA engineers have feedforward control over the spacecraft, because they attempt to calculate and foresee most of the eventualities that could affect the trajectory and the functioning of the spacecraft. The astronaut, by contrast, has feedback control over the trajectory: if something were to divert the spacecraft from its trajectory, she would notice it thanks to the radar and would make proper corrections (the same would happen in case of little breakdowns). Finally, doctors have both feedback and feedforward control over the health of their patients. For example, they have feedback control when they detect new symptoms and provide proper treatment. By contrast, they have feedforward control when they predict and avoid the secondary effects of some medicine that the patient has not yet ingested or when they apply preventive treatments.

As the examples show, monitoring can take many different forms. Accordingly, the dispositions, competences or abilities responsible for monitoring might be very diverse. NASA engineers make mathematical calculations to predict future events; doctors observe, touch and run tests on patients; the heat of the environment drives a thermometer's mercury column long away up the tube; excitement increases the heart rate before exercise above resting levels; bats use echolocation to keep track of the insects they hunt. There is no such a thing as *the* competence or ability to monitor. Different dispositions, competences and abilities might play such a *functional role*. In some cases, those dispositions might be the same as the dispositions responsible for effective control.

§ *Control systems without monitoring.* Unlike feedback and feedforward systems, there are simple systems that do not present any form of tracking control. In particular, they can neither prevent or react to eventual disturbances nor regulate their own activity. The only thing they can do to achieve their goals is to exert effective control (sustained causal influence) over the relevant parameters. In engineering, they are commonly known as *open-loop* systems.

The best way of seeing the difference between open-loop, feedback and feedforward systems is by means of an example. Consider an irrigation sprinkler system. In an open-loop irrigation sprinkler system, the irrigation activates when programmed no matter whether it rains or not. By contrast, a feedback irrigation sprinkler system includes a device for detecting humidity so that it would deactivate the irrigation if it started raining. Finally, a feedforward irrigation sprinkler system is equipped with a device for predicting rain so that it would refrain from activating the irrigation if the device predicted rain. The following diagram illustrates the structure of the three types of control systems:[9]



3.2.5 *Ordinary Control*

What do we mean when we say that *A* controls *B* in an ordinary sense? The term 'control', when used in ordinary contexts, may refer (1) to effective control, (2) to tracking control or (3) to both of them. Let me give some examples:

(1) People typically say that the heart controls blood circulation. On reflection, the kind of control that the heart has over blood circulation is only effective control. Hearts cause blood to circulate but they cannot monitor the blood flow; hormones and medical equipment can.

9 Adapted from Hopgood (2012, section 14.2).

(2) People typically say that thermometers control the environment's temperature. On reflection, thermometers do not control the environment by exerting causal influence over it. Rather, the environment causes changes on thermometers, which enables them to register the temperature (they carry out merely informational monitoring). In this way, ordinary ascriptions such as 'The thermometer controls the temperature' are true in virtue of this kind of tracking control. In addition, ordinary ascriptions of control might be also true when a controller does not exert effective control over a controllee (i.e., when the controller does not drive the controllee to some desired state) and the kind of monitoring carried out by the controller is not merely informational but also dispositional. A good example is a doctor who measures the blood pressure of her patient with a sphygmomanometer. The information puts the doctor in a position to apply proper treatment in case needed (i.e., the information disposes the doctor to take proper action). Ordinary ascriptions such as 'The doctor controls the blood pressure of her patient' are true in virtue of this form of tracking control.

(3) People typically say things such as 'Doctors control the health of their patients'. By reflecting on what people mean when they make such claims, one notices that they sometimes refer both to the fact that doctors monitor that certain parameters are stable (e.g., the blood pressure) and to the fact that they drive those parameters to desired levels (e.g., using stabilizer pills). In other words, ascriptions of control of that type are sometimes true in virtue of both tracking and effective control.[10]

Further characteristics of the notion of control, as we ordinarily understand it, are the following. 1) It is *gradual*: *A* can have more control over *B* than *C* has (e.g., professional soccer players have more control over the ball than amateurs). 2) Control can be *complete*, if *A* has control over all the parameters that are required to control *B*, or *partial*, if *A* has control only over some of those parameters. 3) Control can be *provisional*, if *A* controls *B* for a short period, or *permanent*, if *A* sustains control over *B* for a long period (e.g., institutional control tends to be permanent, as the control that the prison system has over prisoners). 4) Control can be *direct*, if *A* directly intervenes on *B*, or it can be *indirect*, if *A* controls *B* by controlling something else.

3.2.6 *Control and Risk*

Both luck and fortune arise when some event puts an agent at risk. I have distinguished two forms of control that are entailed by our ordinary notion of control. What form of control must fail in order for *S* to be at risk with respect to some event *E*? *S* might be at risk with

10 These points will be very useful to define the notion of achievement in chapter 6.

respect to *E* if *E* is sufficiently significant for *S* and (1) *S* has neither effective nor tracking control over *E*, (2) *S* has tracking control but lacks effective control, (3) *S* has effective control but lacks tracking control.

On the other hand, the amount of effective and/or tracking control that an agent must have to be safe with respect to an event varies from circumstances to circumstances. For any circumstances *C* and any event *E*, there is a minimum level of control that one must reach in order for one to be safe with respect to *E* in *C*. Not all circumstances and events are equally demanding. In the case of effective control this is quite obvious: the level of control that one has to exert in order to drive some item *X* to some desired state depends on the kind of item *X* is and the kind of circumstances in which one is trying to control it. Similar considerations apply to tracking control. The difficulty of monitoring certain parameters depends on the kind of parameters and the circumstances in which one is trying to monitor them. For example, in stable circumstances the causal patterns of which one can avail oneself are always the same. In changing circumstances, however, the relevant causal chains constantly vary, so it is harder to monitor them. Accordingly, circumstances in which there are many changes tend to require more tracking control to be safe from risky eventualities than circumstances in which the relevant parameters are stable.[11] By way of illustration, it is easier to control a sailboat in a region of the sea where the winds are constant than in a region where the winds are constantly shifting.

In addition, even if a causal pattern is favorable to one's interests, one might be at risk if one does not monitor it. How can that be? Does not the fact that the occurrence of an event *E* is favorable to *S* necessarily entail that *E* cannot put *S* in a risky position? No, because if *S* does not monitor *E*, *S* will act under the assumption that *E* is not the case, and here is where risk emerges. Let me give an example:

> Desperate Father
>
> John's daughter is seriously ill and needs a kidney transplant urgently. Sadly, John cannot afford the operation. Fortunately, the government has the policy of financing surgery to children (no matter whether their parents ask for it or not). However, John does not know anything about the government's policies, since he is new in the country. Desperate as he is, he decides to sell one of his organs on the black market to get the money that his daughter needs. John sells one of his kidneys.

11 In philosophy, this sort of circumstances are typically described as 'unfriendly' or as 'non-cooperative'. These terms imply that the environment has goals. However, as we have seen, the environment has no goals. Changing circumstances are better understood as circumstances in which there is risk that things turn out to be different (event-focused sense of risk).

There is no risk that the government does not pay the operation (event-focused sense of risk). However, this favorable event does not put John in a safe position, because John does not know that such a policy exists. Consequently, he acts under the assumption that he has no money, which has very bad effects on his interests. Note that since the government's policy is favorable to John, John does not need to have effective control over the funding of the operation. All he needs is to know that such a health policy exists so that he can *safely rely on* the government to heal his daughter (I will explore the link between controlling X and relying on X in the next section).

§ The relation between event-focused and agent-focused risk. According to Pritchard (2005: 130), although an individual's lack of control over an event (i.e., her being at risk with respect to that event) is a good determinant of whether the event is lucky for the individual, only the modal fragility of the event (i.e., its failing to obtain in close possible worlds) is constitutive of the luck relation between the individual and the event. Riggs (2007) holds the opposite: only lack of control is constitutive of the luck relation. What is the relation between modal fragility and lack of control? Does control entail modal robustness? And *vice versa*?

First, the risk that an event has of not occurring (its failing to occur in close possible worlds) does not necessarily entail the risk that an agent has with respect to the event (her lack of control over the event). Recall whim decisions. If I make a decision on a whim, the decision would not occur in most close possible worlds. However, I definitely have control over the decision. After all, it is *my* decision, self-consciously made. Therefore, control does not entail modal robustness (modal fragility does not entail lack of control). Nevertheless, in most cases where one has control over E, the occurrence of E is safe (modally robust).

Second, the absence of event-focused risk does not entail the absence of agent-focused risk. In other words, the fact that an event is modally robust does not entail that it is under the control of someone. A professional killer may have a strong desire to kill me. Suppose she kills me. Since she was decided to do it and she is very professional, I would have died in all close possible worlds. Yet, my assassination is something over which I have no control, as I do not even know that I have an enemy. Therefore, modal robustness does not entail control (lack of control does not entail modal fragility). In sum, the lack of control and the chance conditions for luck are independent of each other.

3.2.7 *Control and Reliance*

Sometimes, to have control over X we need to rely on events, causal patterns, artifacts and agents to drive X into the state we want X

to be. Reliance poses a problem because it seems that in doing so we lose part of our control over *X*. However, contrary to what one might initially think, not in all cases in which we rely on something or someone we lose control. Well-grounded reliance does not diminish our control but makes it possible and even enhances it.

If we rely on events, causal patterns, artifacts and agents it is because our capacities to exert effective control are limited: in order to control *X* we simply cannot have control over all the factors that causally affect *X*. Nevertheless, we cannot simply place our reliance in those items without any guarantee: our reliance must be such that it does not put us at risk. What is required in order for *S* to rely safely on *X*? At a minimum:

- If *S* does not monitor *X*, it is risky for *S* to rely on *X*.

That is, monitoring is necessary for safe reliance. A caveat is in order, though. The kind of monitoring required cannot only be merely informational monitoring. Registering relevant information about *X* is obviously necessary, but more is needed*:* *S* also needs to monitor *X* in the dispositional sense, i.e., in the sense that the information gathered by *S* puts her in a position or disposes her to take proper action if needed. With this in mind, the following principle plausibly holds as well:

- If *S* relies on *X* and *S* monitors (or has monitored) *X* (not only in the merely informational but also in the dispositional sense of monitoring), *S* safely relies on *X*.

Consider the following example. Marissa, CEO of a big company, safely relies on a subordinate to perform a very technical task that she cannot perform by herself. She can safely rely on her subordinate because she has relevant information about how the subordinate typically works and is in a position to take proper action if needed (e.g., she can hire someone else to perform that task if the subordinate fails).

§ *Trust vs. reliance.* There is a distinction between trust and mere reliance. In the rich literature on trust, philosophers commonly hold, following the work of Annette Baier (e.g., 1986), that the difference between trust and mere reliance consists in the fact that trust can be betrayed whereas reliance cannot. For example, one feels betrayed by a friend if she reveals a secret she promised to keep, but one does not feel betrayed by one's enemy if she unexpectedly attacks one (or by one's car if it breaks down in the middle of the desert). One only feels disappointed. On the other hand, while we can only place our trust in agents, we can place our reliance in agents, artifacts and, in general, on stable causal patterns (e.g., the sunrise, a meteor's trajectory, and so on). In addition, philosophers working on the notion of trust typically contend that trust implies a certain amount of risk or vulnerability whereas reliance does not necessarily so. We can now explain why: when we properly monitor the things we rely on we are at no risk with respect to them. In contrast, when people are incapable of developing even the most basic forms of monitoring, they tend to trust others blindly. For example, think about faith healers and highly superstitious people. Even people with good monitoring skills sometimes cease to monitor certain people and place trust in them (albeit not so blindly). For example, we trust our parents, our partners and our friends not because we monitor all of their actions but because they are our parents, our partners and our friends. Sometimes, well-grounded forms of trust involve low degree of risk, which may suffice for not losing control in certain contexts. However, as a general rule, if you want to safely rely on someone, place no trust in that person: just use your monitoring skills.

3.2.8 *Standards of Evaluation*

We have seen that in control systems the dispositions responsible for tracking control are integrated with the dispositions responsible for effective control in such a way that they establish relations of feedback and feedforward. Thanks to those dispositions, control systems achieve their goals by controlling certain parameters.[12] The question is: what properties should those dispositions have to make the system to which they belong a *good* control system?

To answer this question we can resort to the epistemological literature, where a similar question is made: what properties should the cognitive dispositions of a belief-forming system have in order for it to be a good belief-forming system? Alvin I. Goldman (1986) lists the following three standards: *reliability*, *power* and *speed*. Christopher Lepock (2011) extends the list with these other two: *portability* and *significance-conduciveness*. According to them, these are properties of belief-forming processes, i.e., of psychological processes that have as their main goal the formation of true beliefs. But these properties are applicable to any control system that aims to achieve some goal:

1 *Reliability.* A system is reliable concerning a goal if and only if it has a tendency or a propensity to achieve it and the proportion of achievements of the goal meets a threshold or standard. The threshold is different depending on the type of system and the type of

12 Possessing the ability, competence or, in general, the disposition to drive *B* into the state *A* wants is necessary in order for *A* to have effective control over *B*. For instance, when one controls one's car one must have the ability to move the car into whichever state one wants. Tracking control, by contrast, does not necessarily require ability, competence or disposition to monitor. Monitoring might take place by means other than ability. For example, an agent might be incapable of detecting lies but, on a given occasion, she might detect that her interlocutor is lying thanks to a Good Samaritan, who tells her that her interlocutor is lying. However, for the most part, monitoring arises out of ability, especially in control systems.

situation in which the system is used. For example, a professional striker is reliable if she scores half of the times she shoots at goal. A calculator is not reliable if it delivers the correct result only half of the times. On the other hand, a child might be a reliable striker if she scores one out of ten times.

2 *Power.* A system is powerful concerning a goal if and only if it has a tendency or a propensity to maximize the times it achieves it. As Goldman draws the distinction, reliability is a matter of avoiding failure. Thus, a system can be very reliable if it has a conservative policy: by doing nothing, it avoids error. Power, by contrast, is a matter of maximizing success no matter the number of attempts. In this way, power can have a great cost: by achieving a great number of successes, the system might collect many failures (and reliability might thus decrease). An example: a striker is powerful if she scores many goals even if she fails most of the times she shoots at goal.

3 *Speed.* In order to have control over something, it is very important the speed with which the relevant goal is achieved. For example, strikers usually need to shoot fast to score goals and cooks need to cook fast their dishes to please their clients. In general, fast control systems are preferable to slow control systems (this is more evident in the case of technological systems).

4 *Portability*. Lepock borrows this term from Andy Clark (1997). A system is portable concerning a goal if and only if it can achieve the goal in a wide range of environments. Thermostats, for example, might be very portable because they are able to regulate the temperature in a wide range of (closed) environments. In addition, good drivers are able to drive very different types of cars under very different climate conditions.

5 *Significance-conduciveness.* A system is significant-conducive if it has a tendency or a propensity to achieve goals that satisfy (or contribute to satisfying) important interests of the system. This property is applicable only to living systems as they are the only bearers of interests. Lepock gives an excellent example: primates and humans have innate processes that identify snakes and spiders as dangerous. This innate tendency can be very significant-conducive in certain environments.

These are the kind of properties that make a control system a good control system. Depending on the case, we will be interested in one property or another. Sometimes we are interested in speed (think about automatic trading systems). Other times it does not matter that the system is slow as long as it is reliable (think about a slow but very reliable Russian truck). Sometimes we are interested in portability (think about laptops). In sum, the attributes that make a control system a good control system might vary from case to case.

3.3 LACKEY AGAINST THE LACK OF CONTROL ACCOUNT OF LUCK

In this section, I will focus on a series of counterexamples that Jennifer Lackey (2008) has proposed to the view that

- *E* is lucky for *S* if only if *S* lacks control over *E*,

or, as Lackey puts it, to the view that *E* is lucky for *S* if and only if the occurrence of *E* is significantly beyond *S*'s control. That is, Lackey argues against the view that lucky events are simply events out of control. In my view, lack of control is just necessary for luck. It is precise, therefore, to reply to Lackey's counterexample to the left-to-right direction. Nevertheless, I will also consider Lackey's counterexamples to the right-to-left direction because they will shed some light on the distinction between luck and fortune.

3.3.1 *Counterexamples to the right-to-left direction*

Lackey proposes the following case:

> BAGEL
>
> I walk into my kitchen, toast a bagel, and eat it with cream cheese. Then my husband comes home ten minutes later, my eating a toasted bagel with cream cheese ten minutes earlier is an event that he neither had control over (he wasn't home) nor was sufficiently responsible for (he had nothing to do with my eating the bagel in question). But is it lucky for him that I ate a toasted bagel with cream cheese? If so, it is clearly not in any interesting sense of luck. (Lackey 2008: 256-257)

Other mundane counterexamples presented by Lackey are: one's neighbor's playing a computer game right now, one's cat sleeping this afternoon or a chef's making eggplant parmesan in Florence today. Lackey is completely right in thinking that these cases do not count as cases of luck in any interesting sense of the term. The reason is simple: the events in question are not sufficiently significant and hence not lucky. In other words, the condition that fails in those cases is the significance condition. Nonetheless, Lackey anticipates this kind of reply and proposes similar cases involving this time significant events:[13]

13 Curiously, Lackey transforms BAGEL into a clear-cut case of luck by keeping all the features of the case fixed except for the significance of the event, which is increased: "suppose that my husband's health requires that he be gluten-free, but he is nonetheless occasionally overcome with powerful cravings for bread and related food items. My eating a toasted bagel with cream cheese, then—which happened to be the last item of food in our house that contained gluten—removed a temptation from his environment that he would not have been able to resist, thereby saving him from a debilitating physical reaction" (Lackey 2008: 257).

Isabella

> My picking up my 6-year-old daughter, Isabella, from school while my husband is teaching is an event that is not only outside of his control, it is also one that is deeply significant and important to him. But surely it is not a lucky event that Isabella is picked up from school. I pick her up from school every day at the same time; I have never forgotten her, nor have I ever arrived to her school late; I am not the sort of person who would neglect my commitments where my children are concerned; and so on. (Lackey 2008: 257-258)

Catherine

> My younger daughter, Catherine, who is 4 years old, is well-fed, clean, and safe on a daily basis. Given her tender age, however, she is not responsible for many of the events that lead to her being properly cared for, nor is she in control of many of these events, despite the fact that they are very significant and important to her. (Lackey 2008: 258)

Interestingly, Lackey thinks that "there is a sense in which both of the events discussed above are lucky: my husband is lucky that he has the sort of wife whom he can depend on to pick up their children, and Catherine is lucky that she has a father who takes proper care of her" (Lackey 2008: 258). However, she argues that this sense of luck makes us deem too many events as lucky and it clearly differs from the sense of luck of the type of events in which proponents of Lack of Control are interested (e.g., fair lottery wins). The sense of luck to which Lackey is referring is what I have called fortune, whereas proponents of Lack of Control are typically interested in what I have called luck.[14] A lottery win is lucky because one does not control the lottery process (agent-focused sense of risk) and because it could easily have not occurred (event-focused sense of risk). In Isabella and Catherine, there is no risk that the target events do not occur. Therefore, the cases are not cases of luck. Rather, they *seem* cases of fortune.

We cannot straightforwardly affirm that Lackey's examples are clear-cut cases of fortune because it is not obvious that 1) Lackey's husband has no control over the fact that she picks up her daughter Isabella from school and that 2) Catherine has no control over her being properly cared for. Consider the case of Catherine. Although she is not directly responsible for many of the events that lead to her being properly cared for, Catherine knows that she can count on her mother

14 Remember that in ordinary discourse, we use the words 'luck' and 'fortune' interchangeably. My use of the terms 'luck' and 'fortune' corresponds to the definitions given in chapter 2.

for being cared for. In the same way, Lackey's husband is certainly confident that her wife will pick up his daughter from school. In both cases, even though the agents in question do not have any direct causal influence over the events described (i.e., even though they do not have (sufficient) effective control over the relevant actions), they certainly *rely* on Lackey for doing them. Depending on whether the reliance placed is well-grounded or not and, in particular, depending on whether they are able to monitor Lackey's actions, we would say that they have control or not over the target events.

To compare, imagine a case in which a child knows of the existence of social services and calls for aid to remedy the carelessness of her parents. Is the child fortunate for being cared for? It does not seem so. The child cannot manage on her own, true, but she can keep track of the attention received and knows how to resort to someone who can take care of her. In the same way, the events that lead to Catherine being properly cared for are not fortunate for her if she can keep track of the attention received and if she knows to whom she can resort or what to do in case of serious domestic problems (a newborn, for example, cannot not do such things). In the same way, Lackey's husband is not fortunate for the fact that his daughter is picked up from school safely if, say, he can call his wife at any time to know whether she is on her way to Isabella's school.

3.3.2 *Counterexample to the left-to-right direction*

Lackey's alleged counterexample to the necessity of Lack of Control for luck is the following:

> Demolition Worker
>
> Ramona is a demolition worker, about to press a button that will blow up an old abandoned warehouse, thereby completing a project that she and her co-workers have been working on for several weeks. Unbeknownst to her, however, a mouse had chewed through the relevant wires in the construction office an hour earlier, severing the connection between the button and the explosives. But as Ramona is about to press the button, her co-worker hangs his jacket on a nail in the precise location of the severed wires, which radically deviates from his usual routine of hanging his clothes in the office closet. As it happens, the hanger on which the jacket is hanging is made of metal, and it enables the electrical current to pass through the damaged wires just as Ramona presses the button and demolishes the warehouse. (Lackey 2008: 258)

Lackey claims that the explosion is both under Ramona's control and by luck. However, not everyone will be willing to accept one of these two intuitions. In what follows, I will explain two ways of diminishing the force of the counterexample, which will show, in turn, why the intuitions of the case are unstable. On the one hand, we can grant the intuition that the explosion is by luck and dispute the assumption that Ramona has control over the explosion. On the other hand, we can grant the intuition that there is control and dispute the assumption that the coincidence described makes the explosion lucky. Let us see these two options in more detail.

1. *Uncontrolled lucky explosion.* As the case is described, there is risk that the explosion does not take place (event-focused sense of risk). However, as we have seen in chapter 1, risk of nonoccurrence does not suffice for luck: there must be risk *for* the agent that the event in question does not occur. Given our conception of agential risk, the question of whether Ramona is at risk with respect to an eventual failure of the explosion amounts to the question of whether Ramona has control over the explosion.[15]

What does it take to have control over an explosion? In particular, does the merely pressing of the button suffice for having control over the explosion? It does not seem so, since an important feature of the case, as Lackey presents it, is that Ramona is one of the persons who have been working for weeks on the design of the controlled explosion. In this way, the extent to which Ramona should have control over the explosion does not only encompass the mere production of the explosion by pressing the button, but also the monitoring of the explosion system.

Therefore, not having properly checked the relevant wires before the explosion or having failed to foresee the presence of rodents, termites or other problematic animals are things for which we can blame Ramona and the rest of the people responsible for the design of the alleged controlled explosion. If Ramona or her co-workers had monitored these things, they would have been in a position to take proper action and thus to demolish the warehouse in a way that we would not classify as lucky. Therefore, while it is true that Ramona causes the explosion and, in a sense, she controls the explosion (effective control), she does not monitor the parameters that are relevant to the explosion (tracking control). Since she lacks overall control over the explosion, the explosion occurs by luck.

2. *Controlled non-lucky explosion.* The second way of diminishing the force of the counterexample grants Lackey's assumption that Ramona has control over the explosion just by pressing the button, but disputes that the explosion is by luck. Levy (2009, 2011) and Coffman (2009) follow this strategy. Coffman, in particular, argues that Lackey,

15 An eventual failure of explosion would be detrimental to Ramona's subjective interests, particularly, to her job.

by claiming that Ramona is lucky with respect to the explosion, upholds what he calls the *luck infection thesis*, the thesis that if *S* was lucky to be in a position to φ, *S* was lucky to have φ-ed.

This thesis, however, has some blatant counterexamples. God may decide, by flipping a coin, to cure my impaired vision, but once I can see it is not by luck that I come to know that my hands are stained by black ink. An allied experienced sniper may be lucky to have Hitler within the range of his rifle (suppose that both the sniper and Hitler are there by sheer coincidence). However, given his expertise, it is not by luck that the sniper shoots Hitler in the head. In sum, it is not the same thing to be lucky in being positioned to φ than being lucky in φ-ing.

In Demolition Worker, Ramona's co-worker hangs his jacket on a nail in the precise location that *enables* the electrical current to pass through the damaged wires. On this view, this fact does not make the explosion lucky, but luckily puts Ramona in a position to activate the explosive charges in such a way that, once she is in that position, it is not by luck that the warehouse is demolished.

The mouse acts like a fink causing the explosion system to lose its disposition to demolish the warehouse when the button is pressed. The hanging of the jacket by the co-worker in the right place at the right time allows the system to regain that disposition. Even if the disposition could easily have been lost (by removing the jacket), the demolition of the warehouse is creditable to the system that Ramona and her co-workers have designed. That is, if the explosive charges are activated when Ramona presses the button it is because the explosion system still has (albeit luckily) its original capacity to demolish the building. As Levy remarks, "[a]ctions which rely upon luckily satisfied causal circumstances do not inherit that luck from the circumstances" (Levy 2009: 492).

In conclusion, Demolition Worker prompts unstable intuitions. This is its trick. For some, it will not be so obvious that Ramona has control over the explosion (this is my position). For others, it will be dubious that the explosion is by luck. Lackey presents the case in a way that makes us think that the two intuitions come together but, on reflection, they do not: the two intuitions come apart. Therefore, we can safely conclude that Lack of Control is a necessary condition for both luck and fortune.

3.4 SUMMARY

In this chapter, I have analyzed the lack of control condition for luck and fortune: the condition that an event is lucky/fortunate for an agent only if the agent lacks control over the event. In section 3.1, I have analyzed E. J. Coffman's version of the lack of control condition in terms of lack of choice. I have shown that one might be lucky even though one had the choice to make something that would produce or prevent the relevant lucky event, because one can deliberately refuse to exercise one's freedom. In section 3.2, I have advanced a general account of the notion of control. I have distinguished two types of control: effective and tracking control. Roughly, *A* has effective control over *B* just in case *A* drives *B* into the state *A* wants. *A* has tracking control over *B* just in case *A* monitors *B.* I have explained that our ordinary ascriptions of control might be true when *A* has either effective control or tracking control or both forms of control over *B*. In addition, I have explored the relation between the agent-focused and the event-focused senses of risk and the relation between control and reliance. Finally, I have reviewed some of the properties that make a control system a good control system. In section 3.3, I have defended the lack of control condition from a counterexample given by Jennifer Lackey. The first stage of the project has been thus accomplished: to give a distinctive and complete account of the concept to which the terms 'luck' and 'fortune' refer when ordinarily used. The next stage is to specify the sense in which knowledge is incompatible with luck. This is the topic of the next chapter.

Part II

KNOWLEDGE

4 EPISTEMIC LUCK AND EPISTEMIC FORTUNE

4.1 UNGER/PRITCHARD'S TAXONOMY OF EPISTEMIC LUCK

Peter Unger (1968) distinguishes several accidents concerning factors that enable or cause the acquisition of propositional knowledge. In his 2005 monograph *Epistemic Luck*, Duncan Pritchard picks up the thread of the discussion and classifies them in a taxonomy of types of epistemic luck (or, as Unger would put it, of epistemic accidents), which shows, among other things, that not all varieties of epistemic luck are incompatible with knowledge. Any analysis of this important epistemic phenomenon should seriously take into consideration Unger/Pritchard's taxonomy. It is therefore my aim in this section to analyze it.[1]

§ *Epistemic factors*. The following list encompasses several factors that enable or cause the acquisition of propositional knowledge. For a proposition *p* to be known by an agent *S* it is necessary that:

1. *S exists*.
2. *S* possesses the relevant *cognitive abilities*.
3. *S forms* the belief that *p*.
4. There is a *truth-maker* for *p* / the proposition *p* is true.
5. *S* has certain *evidence* in favor of the belief that *p*.
6. *S comes to believe the truth* about *p*.[2]

1 Unless otherwise indicated all references to Unger and Pritchard will be to Unger (1968) and Pritchard (2005).

2 Note that (3), (4) and (6) are different necessary factors for knowing a proposition *p*. (3) concerns the fact that the agent forms the belief that *p*, i.e., the fact that she possesses the doxastic state with content *p*, regardless of whether that content is true or false. (4) concerns the fact that in the world there is a truth-maker for the proposition *p* (regardless of the doxastic states that have as content such a proposition). (6) concerns the fact that the agent's belief, so to speak, 'hits' the truth about *p*. A precondition for (6) is obviously (3) and (4). In other words, (6) can hold only if (3) and (4) are in place. In addition, if (3) and (4) obtain, then (6) is the case. That is, *S comes to believe the truth* about *p* if and only if *S forms* the belief that *p* and there is a *truth-maker* for *p*. Although (given this equivalence) it might be redundant to introduce (6) as a distinctive necessary factor for possessing propositional knowledge, it is useful to do so because it will allow to identify a distinctive type of epistemic luck.

§ *Types of epistemic luck*. As Unger noted, his analysis of knowledge in terms of lack of accidentality radically departed from all analyses offered to the date, 1968 (i.e., five years after the publication of Gettier's seminal paper "Is Justified True Belief Knowledge?"), as it did not focus on modifying the justification condition of the tripartite analysis of knowledge or on supplementing it with other epistemic conditions. Although Unger's analysis of knowledge (*S* knows that *p* if and only if it is not accidental that *S* is right about *p*) was vague and not very illuminating, he motivated it with several examples that aimed to show not only that factors 1-6 may obtain by accident, but also, and more importantly, that not all epistemic accidents undermine knowledge. Unger's important contribution to the literature was not therefore his definition of the concept of knowledge but the distinction of several epistemic accidents that constitute, according to Pritchard, different varieties of epistemic luck.[3] To begin with, factors 1 and 2 may be affected by what Pritchard calls:

- *Capacity epistemic luck*: it is lucky that the agent is *capable* of knowledge.

The formulation of capacity luck is intended to cover both the luck that an agent has to exist (luck concerning factor 1) and the luck that an agent has to possess the psychological and physiological constitution required for knowledge (luck concerning factor 2). Further varieties of epistemic luck are:

- *Doxastic epistemic luck*: it is lucky that the agent *believes* the proposition. (Factor 3)

- *Content epistemic luck*: it is lucky that there is a *truth-maker* for the proposition, i.e., it is lucky that the proposition is true. (Factor 4)

- *Evidential epistemic luck*: it is lucky that the agent acquires the *evidence* that she has in favor of her belief. (Factor 5)

- *Veritic epistemic luck*: it is a matter of luck that the agent *comes to believe p truly*. (Factor 6)[4]

§ *Examples*. The following series of examples illustrates the different forms of epistemic luck:

A. *Capacity luck example (factor 1)*: In the beginning of time, God flips a coin to decide whether to create Adam or Peter. The coin

3 The credit for the taxonomy of epistemic luck is therefore partly Pritchard's and partly Unger's.

4 Pritchard (2004; 2005) borrows the expressions 'evidential epistemic luck' and 'veritic epistemic luck' from Engel (1992), but they use them differently. While Engel explicitly mentions the agent's evidence in his definitions of evidential and veritic luck, Pritchard only points to it in his definition of evidential luck.

lands heads and God creates Adam. Subsequently, Adam sees that he has two hands and forms the true belief that he has two hands. Intuitively, Adam knows that he has two hands.

B. *Capacity luck example (factor 2)*: Subsequently, God flips a coin to decide whether to endow Adam with a reliable auditory system. The coin lands heads and God endows Adam with a reliable auditory system. Adam speaks, hears his voice for the first time and comes to believe that he has a deep voice. Intuitively, Adam knows that he has a deep voice.

C. *Doxastic luck example*: There are no mirrors in Eden, and Adam has never seen his back, and therefore he does not know the number of moles on his back. God hypnotizes Adam and flips a coin to decide whether to induce in Adam the true belief that he has ten moles on his back or the false belief that he has eleven moles. The coin lands heads and God induces in Adam the true belief that he has ten moles on his back. A short while later, Adam wakes up and truly believes that he has ten moles on his back, but, intuitively, he does not know it.

D. *Content luck example*: God flips a coin to decide whether to endow Adam with knees or not. The coin lands heads and God endows Adam with knees. Subsequently, Adam takes a look at his legs and forms the true belief that he has two knees. Intuitively, Adam knows that he has two knees.

E. *Evidential luck example*: Adam has been expelled from Eden and becomes a bank robber (suppose that everything that is not Eden is Hell and that Hell is full of bankers). As he is escaping from the bank that he has just robbed, his mask slips off for a few seconds allowing a police officer to identify him. Intuitively, the police officer knows that Adam has robbed the bank.[5]

F. *Veritic luck example*: God allows Adam to go back to Eden. Adam sees a snake-looking thing in a tree. The thing Adam is looking at is in reality a snake-looking branch. When he is about to form the belief that there is a snake in the tree, a snake fortuitously climbs the unseen side of the trunk, so that Adam's belief is true. Intuitively, Adam does not know that there is snake in the tree.[6]

§ *Harmful vs. harmless epistemic luck*. In the literature, it is not uncommon to come across the claim that luck is incompatible with knowledge. In the light of what is perhaps the most important of Unger's

5 Case originally introduced by Nozick (1981: 93).
6 Example inspired by Chisholm's Gettier-style sheep-in-the-field case (Chisholm 1977: 105).

findings (that an agent can know a proposition when some of the factors that enable or cause that instance of knowledge obtains by sheer luck), the claim obviously needs qualification. In fact, as we will see next, most forms of epistemic luck can coexist with knowledge. Let us take a look at the data to identify epistemically harmless forms of luck.

Cases A, B, D and E are cases of knowledge. Therefore, *capacity luck* (cases A and B), *content luck* (case D) and *evidential luck* (case E) are compatible with knowledge. By contrast, we will hardly find a case in which and agent comes to believe the truth by luck and the agent knows. As case F exemplifies, there is good reason to think that *veritic luck* is incompatible with knowledge.[7] Is *doxastic luck* compatible with knowledge possession? Pritchard thinks it is. After all, although case C, which is a case of doxastic luck, is not a case of knowledge, case E, which is a case of knowledge as well, not only instantiates evidential luck but also doxastic luck. In case E, the police officer gathers by luck evidence that allows her to identify Adam. If it is a matter of luck that she acquires that evidence, then it is also a matter of luck that she forms the belief that Adam has robbed the bank, something that she knows. Doxastic luck is, therefore, compatible with knowledge.

4.1.1 *Veritic Luck*

Let us take a closer look to veritic epistemic luck. Pritchard's definition of veritic luck is an adaptation of his modal chance condition for luck (MODAL CHANCE 1)[8] to cases in which the relevant epistemic event is an agent coming to believe a proposition truly:

- VERITIC LUCK$_{Pritchard}$: *S*'s true belief is [veritically] lucky iff there is a wide class of near-by possible worlds in which *S* continues to believe the target proposition, and the relevant initial conditions for the formation of that belief are the same as in the actual world, and yet the belief is false. (Pritchard 2007: 280)

Case F is a paradigmatic case of veritic luck. In most close possible worlds in which Adam forms the belief that there is a snake in the tree, there is no snake in the tree (remember that the snake fortuitously climbs it). Plausibly, this is the reason why Adam's actual true belief is not knowledge. Can we conclude, in the light of this sketchy diagnosis, that veritic luck is incompatible with knowledge? To draw such a conclusion, we need to explain first why the following

7 Case F exemplifies well the difference between doxastic and veritic luck: it is not by luck that Adam forms the belief that there is a snake in the tree, but it is certainly by luck that he comes to believe the truth.

8 MODAL CHANCE 1 says that an event *E* is lucky for *S* only if *E* occurs in the actual world but does not occur in a wide class of the nearest possible worlds where the relevant initial conditions for *E* are the same as in the actual world (Pritchard 2005: 128).

case of *knowledge*, which is reasonably similar to case F (at least in its structure), is *not* an instance of veritic luck:

> GRANDMOTHER
>
> A grandmother goes to the hospital to visit her grandson. The grandson's heart is in such a very bad condition that he could easily suffer a deathly cardiac arrest at any moment. If the grandson died, others would tell the grandmother he was alive to spare her upset. The grandmother enters the hospital room and sees that the grandson is alive.[9]

Intuitively, the grandmother knows that her grandson is alive when she enters the hospital room and sees him awake and breathing. In addition, in most close possible worlds in which she forms the same belief, her grandson is dead and she falsely believes that he is alive (others tell her so). Is then her actual true belief veritically lucky? Is, therefore, veritic luck compatible with knowledge? Such a conclusion overlooks one fundamental aspect of the notion of veritic luck as Pritchard defines it, one aspect that allows to explain why case F, but not GRANDMOTHER, is a case of veritic luck.

According to VERITIC LUCK$_{Pritchard}$, *S*'s true belief that *p* is luckily true just in case (1) *p* is false in most close possible worlds in which *S* forms the belief that *p* and (2) in which the relevant initial conditions for the formation of that belief are the *same* as in the actual world. What does it count as the relevant initial conditions for the formation of a belief? Many factors might be relevant for the formation of a belief, but the most important factor is obviously the method of belief formation used to formed the belief. GRANDMOTHER *seems* to prove that veritic luck is compatible with knowledge because (1) holds: it is not the case that the grandson is alive in most close possible worlds. However, (2) does not hold. In the actual world, the grandmother comes to believe that her grandson is alive by seeing him alive (her method of belief formation is visual). Close possible worlds in which the grandmother believes falsely that her grandson is alive (most close possible worlds) are worlds in which she forms her belief by testimony.

Of course, there are few close possible worlds in which the grandmother believes that her grandson is alive by the same type of visual method and therefore there are few close possible worlds in which the relevant initial conditions for the grandmother's belief are the same as in the actual world, but since in most of those worlds her grandson is still alive, her actual belief is *not* veritically lucky. Compare GRANDMOTHER with case F. Adam's actual belief that there is a snake in the tree is veritically lucky because he does use the same type of method of belief formation in close possible worlds in which

9 Case inspired by Nozick's grandmother case (Nozick 1981: 179).

he believes the same proposition falsely (as in the actual world, he *looks* at the tree). Consequently, the relevant initial conditions for his belief in those worlds are the same as in the actual world. In sum, while in case F (1) *and* (2) hold, in GRANDMOTHER only (1) holds. This asymmetry explains the absence of knowledge in the former case and of veritic luck in the latter.

4.1.1.1 *Veritic Intervening Luck*

When Adam forms his belief that there is a snake in the tree, he is not looking at a snake, but at a snake-looking branch. If a snake had not fortuitously climbed the hidden side of the trunk just when Adam was about to form his belief, he would have formed a false belief. But, luckily, the snake climbs the tree and as a consequence Adam comes to believe the truth.

Pritchard calls *intervening luck* the kind of veritic luck involved in cases structurally equivalent to case F, because in all of them luck 'intervenes' in the way the agent gets things right, in the sense that luck is what 'links' the agent's belief with the truth. What Pritchard has in mind, I think, is (roughly) the following: in cases of veritic intervening luck the agent's cognitive abilities do not explain why she comes to believe the truth in the actual world.

To see this, consider again case F. Adam forms the belief that there is a snake in the tree by looking at a branch that looks like a snake. His visual cognitive abilities do not seem to explain why he comes to believe the truth. He comes to believe the truth because when he is about to form his belief a snake fortuitously climbs the unseen side of the trunk. It is in this explanatory sense in which luck is said to 'intervene'.[10]

To grasp the point better, we can introduce an analogy considered by Pritchard (2009b) and extensively used by virtue epistemologists (mainly by Ernest Sosa) with several theoretical purposes. Suppose that knowers are like archers, their beliefs like arrows and the truth like a bull's eye. A case of intervening luck would be a case in which a skillful archer aims at the target, shoots and a fortuitous gust of wind deviates the arrow from its trajectory but then again another unexpected gust brings it back to its original trajectory thus making the arrow hit the bull's eye. In this case, the type of luck at issue is of the intervening sort because what saliently explains the archer's success is not her mastery of the bow but the fortuitous double intervention of the wind.

10 Another way to explain cases of veritically lucky belief is the following: there is a causal disconnection between the truth-maker of p and the fact that S comes to believe that p truly. Nevertheless, this might not be applicable to all cases, as it is possible that the truth-maker of p causes that S comes to believe that p and, in a deviant way, that S comes to believe that p truly (and luckily).

4.1.1.2 *Veritic Environmental Luck*

Not all cases of veritic luck are like case F. Consider the following much-discussed example:

> FAKE BARNS
>
> Henry is driving in the countryside with his son. They play a game called 'Object Identification'. Henry sees a barn and forms the belief that there is a barn in front of him. The object is a barn in full view. Henry has excellent eyesight, and he has enough time to look at it reasonably carefully, since there is little traffic to distract him. Suppose that, unknown to Henry, the district he has just entered is full of papier-mâché facsimiles of barns. These facsimiles look from the road exactly like barns, but are really just facades. The object Henry sees is a genuine barn, but if the object on that site were a facsimile, Henry would mistake it for a barn.[11]

The object Henry sees is a genuine barn in full view. Henry has excellent eyesight and has enough time to look at it reasonably carefully. On these grounds, Henry forms the belief that the object in front of him is a barn. Accordingly, Henry's exercise of his perceptual abilities explains why he comes to believe the truth. Importantly, when one has propositional knowledge about the presence of barns one typically forms beliefs in this way. This particular feature has led some epistemologists to subscribe that agents have (at least some form of) propositional knowledge in this kind of scenarios (e.g., Sosa 2007: 35). Nevertheless, most epistemologists in the specialized literature have the intuition that Henry's belief is true by luck.[12]

The kind of veritic luck at issue, however, is not intervening luck, since Henry's cognitive abilities *do* explain why he comes to believe that there is a barn in front of him truly. The reason Henry does not know that the object in front of him is a barn is that the *environment* is such that he could easily have looked at a barn replica, which would have induced him to believe the same proposition falsely. It is not a coincidence that Pritchard calls this variety of veritic luck *environmental luck*.

As before, the following is (roughly) what I think Pritchard has in mind: in cases of veritic environmental luck the agent's cognitive abilities explain why she comes to believe the truth in the actual world,

11 The case is Carl Ginet's and has been spread by Goldman (1976).

12 Sosa's ascription of knowledge to Henry (and to other individuals in similar situations) hinges on a distinction between two levels of knowledge: animal and reflective. While Sosa thinks that Henry has animal knowledge, he explains the intuition that there is no knowledge (an intuition that, as I say, is shared by most epistemologists) by appealing to Henry's lack of reflective knowledge (it is not clear whether the difference between the two types of knowledge is a difference of degree or a qualitative difference).

but features of the environment would prevent her from believing the truth in close possible worlds (albeit not in the actual world).

To make more clear the distinction between the two varieties of veritic luck, let us resort again to the archery analogy. In the case described before, shifting winds make the arrow luckily hit the bull's eye and hence the exercise of the archer's skills do not explain why her shot is successful. In the present case winds are normal. The skillful archer takes aim, shots and hits the bull's eye. Everything is alright, except for the fact that she has randomly shot at the only target among thousand of possible targets that is not protected with an invisible forcefield that would have repelled any incoming arrow. The archer's success is thus explained by her shooting skills, but the environment is such that she could easily have missed the shot. Therefore, the type of luck at issue is of the environmental sort.

4.1.2 *Reflective Luck*

Pritchard distinguishes another potentially dangerous variety of epistemic luck, which he calls:

> REFLECTIVE LUCK: For all agents, φ, the truth of an agent's belief in a proposition, φ, is reflectively lucky if, and only if, the agent's belief that φ is true in the actual world, but, in nearly all possible worlds consistent with what the agent is able to know by reflection alone, were the agent to form a belief that φ, that belief would be false. (Pritchard 2003: 122)

§ *Points of clarification*. We will consider some examples of reflective luck in a moment. Before that, two points must be highlighted in order to understand Pritchard's definition properly:

1. The world-ordering that is needed to establish whether a belief is reflectively lucky is a non-standard one: possible worlds relevant to the definition above are only those that are consistent with what the agent is able to know by reflection alone in the actual world. As Pritchard insists, this means, first, that the relevant worlds must not be ordered in terms of how the agent in fact formed her belief, but in terms of how the agent believes (or would believe) she formed her belief in the actual world; second, this means that not every second-order belief about how the agent has formed her first-order belief contributes to the world-ordering, but only those that are acquired by reflection alone. Pritchard (2005: 180, fn. 18) suggests that in the particular case that the agent does not or would not form any second-order belief whatsoever, no restriction applies to the relevant range of possible worlds; in other words, any possible world could count as nearby.

2. We need to elucidate exactly what Pritchard means by the term 'reflection'. The answer can be found in a clarification that he makes about his definition of epistemological internalism about justification:[13]

> [B]y 'reflection', I mean a priori reasoning, introspective awareness of one's own mental states, and one's memory of knowledge that has been gained in either of these ways. (Pritchard 2005: 42)

Nevertheless, Pritchard allows for less restrictive interpretations, like the following proposed by Sosa:

> By definition, "reflection" involves either (a) introspection, (b) rational intuition, (c) memory, (d) deduction, or (e) induction or ampliative reason which builds only on materials provided by (a)–(d). (Sosa 2003: 144, fn. 5)

It is clear what does not count as 'reflection': the two traditional sources of knowledge excluded from the lists above, viz., perception and testimony. Accordingly, no second-order belief stemming from these sources will contribute to the reflective world-ordering required to determine whether a true belief is reflectively lucky.

§ *Examples*. Having clarified Pritchard's definition of reflective luck, we move on to consider some examples, which I will group into three different sets. The reason to do so is that each set of cases shows, as we will see, different ways in which reflective luck might be epistemically problematic.

Set 1. The first set includes several cases that have been at the center of the internalist-externalist controversy:

> Chicken-Sexer
>
> A professional chicken-sexer looks at a chick and forms the true belief that it is a male. The chicken-sexer is unaware of the process by which he tells the sex of the chick, but, as always, he is correct in his judgment. (Goldman 1975: 114)

13 Pritchard defines internalism about justification as follows: "For all agents, φ, an agent's belief that φ is justified if, and only if, the agent is able to know the facts which determine that justification by reflection alone" (Pritchard 2005: 42). I will use this conception of internal justification in my discussion of reflective luck. Internalist justification so conceived amounts to *access internalism*, the view that all the factors needed for a belief to be justified must be cognitively accessible to the agent. Although access internalism is the most common characterization of internalism about justification, there are alternative ways of conceiving internalism. *Mentalism*, for example, is the view that agents obtain justification for their beliefs only from factors internal to their mental lives. More precisely, as Earl Conee and Richard Feldman put it, mentalism holds that "the justificatory status of a person's doxastic attitudes strongly supervenes on the person's occurrent and dispositional mental states, events, and conditions" (Conee & Feldman 2004: 56).

Chicken-sexers are real and so were plain spotters:

> PLAIN SPOTTERS
>
> During World War II, under constant threat of bombings, the British had a great need to distinguish incoming aircraft quickly and accurately. Which aircraft were British planes coming home and which were German planes coming to bomb? Several airplane enthusiasts had proved to be excellent "spotters," so the military eagerly employed their services. These spotters were so valuable that the government quickly tried to enlist more spotters but they turned out to be rare and difficult to find. The government therefore tasked the spotters with training others. It was a grim attempt. The spotters tried to explain their strategies but failed. No one got it, not even the spotters themselves. (Eagleman 2011: 58)

The following two well-known cases, proposed by two well-known champions of internalism, have helped to inform the debate between internalists and externalists in epistemology:

> CLAIRVOYANT
>
> Norman, under certain conditions that usually obtain, is a completely reliable clairvoyant with respect to certain kinds of subject matter. He possesses no evidence or reasons of any kind for or against the general possibility of such a cognitive power, or for or against the thesis that he possesses it. One day Norman comes to believe that the President is in New York City, though he has no evidence either for or against this belief. In fact the belief is true and results from his clairvoyant power, under circumstances in which it is completely reliable. (BonJour 1980: 62)[14]
>
> TRUETEMP
>
> Suppose a person, whom we shall name Mr. Truetemp, undergoes brain surgery by an experimental surgeon who invents a small device which is both a very accurate thermometer and a computational device capable of generating thoughts. The device, call it a tempucomp, (...) is very reliable, and so [Truetemp's] thoughts are correct temperature thoughts. All told, this is a reliable belief-forming process. Now imagine, finally, that he has no idea that the tempucomp has been inserted in his brain, is only slightly puzzled about why he thinks so obsessively about the temperature, but never checks a thermometer to determine

14 To make more powerful BonJour's intuitions about the case, suppose that that is the first belief that Norman forms using his clairvoyant powers.

> whether these thoughts about the temperature are correct. He accepts them unreflectively, another effect of the tempucomp. (Lehrer 1990: 163-164)

In this set of cases, all the agents possess unusual cognitive faculties: clairvoyance powers, the capacity to detect the temperature and the abilities to discriminate female from male chicks and British from German planes. None of them is aware of how they form their true beliefs.

Set 2. In the second set of examples of reflective luck, no agent exhibits odd cognitive faculties. On the contrary, the cases describe common situations in which agents employ familiar cognitive abilities to form true beliefs in a way that is not transparent to them.

> BROCCOLI
>
> Last year, Sally read a story about the health benefits of broccoli in the "Science" section of the *New York Times*. She then justifiably formed a belief in broccoli's beneficial effects. She still retains this belief but no longer recalls her original evidential source (and has never encountered either corroborating or undermining sources). (Goldman 1999: 280)
>
> FACE RECOGNITION
>
> Harry and Sally met when both needed someone to share a drive to New York City. Five years later, they are both in a New York airport. Harry asks for a coffee, looks at the woman next to him and comes to believe that the woman next to him is Sally. It is Sally, indeed. However, if asked, Harry could not tell how he has recognized her.[15]

Sally and Harry's cognitive abilities are definitely common. Sally reads a true story from a reliable source and comes to believe a proposition about the benefits of eating broccoli. After some time, she still retains the belief but she cannot recall the original evidential source, which makes her true belief reflectively lucky. The case of Harry is even more mundane. He successfully recognizes Sally's face but he cannot tell how he has carried out such a successful cognitive performance.

Set 3. The distinctive feature of the third and last set of examples of reflective luck is that it only includes skeptical scenarios such as cases of massive deception. Pritchard's definition of reflective luck nicely applies to these cases because massively deceived agents are obviously not reflectively aware of the way they form their beliefs. Consider one of the many examples that populate the literature and our collective imagination:

15 Inspired by Lewis's discussion on easy knowledge (1996: 551).

NEO

> Unbeknownst to Neo, evil machines have connected him to a simulated reality called the Matrix. They feed him experiences of an external world, so that Neo comes to believe certain empirical propositions. He also thinks that he forms his beliefs as he used to: using his perceptual abilities. Some of the beliefs he acquires are true.

All the cases of the three sets are cases of reflective luck, because the agents would not believe the truth in most possible worlds consistent with what they are able to know by reflection alone. Consider CHICKEN-SEXER. First, given that the relevant world-ordering that matters to determine whether the chicken-sexer's beliefs are reflectively lucky is established solely by his second-order beliefs about how he forms his beliefs about chicks in the actual world, and second, given that he is completely unaware of how he discriminates chicks, there is no guarantee that in close possible worlds (non-standardly ordered) he still has his discriminatory ability and, therefore, that he successfully discriminates chicks in those worlds. The same reasoning applies to plain spotters (PLAIN SPOTTERS), Norman (CLAIRVOYANT), Mr. Truetemp (TRUETEMP), Sally (BROCCOLI) and Harry (FACE RECOGNITION). In the skeptical scenario of NEO, we know that Neo falsely believes that he forms his true beliefs about the external world using perceptual abilities. Since his second-order beliefs are false, he fails to believe the truth in close possible worlds ordered in terms of what he is able to know by reflection alone in the actual world and, therefore, his first-order true beliefs are reflectively lucky.

4.1.2.1 *How Problematic Is Reflective Luck?*

The question that immediately comes to mind is whether reflective luck is incompatible with knowledge, and if it is not, how problematic it is. Interestingly, the extent to which reflective luck is considered epistemically dangerous largely depends on our epistemological sympathies and namely, as Pritchard aims to show, on whether they lie with externalism or internalism.

§ Reflective luck and the externalist/internalist divide. Controversy arises when externalists and internalists discuss cases of the first set. Typically, externalists do not find anything wrong with ascribing knowledge (although perhaps not justification) to agents with unusual cognitive abilities such as chicken-sexers, plain spotters, clairvoyants or temperature detectors, provided that their cognitive faculties are reliable. By contrast, according to internalists, these agents are not internally justified in their true beliefs (and fail to know if internal justification is considered necessary for knowledge) because they lack *reflective access* to the facts that help to fix the justificatory status of

their beliefs, that is, to the relevant justifiers (e.g., to the reliability of their cognitive faculties or the reasons in virtue of which they hold their beliefs).

Reflective luck precisely arises if an agent lacks reflectively accessible grounds to back up her beliefs and hence if her beliefs fail to be internally justified. The problem of explaining whether or not reflective luck is compatible with knowledge can be thus momentarily understood as the problem of explaining whether or not lack of internal justification entails lack of knowledge. In what follows, I will consequently focus on the following implication: if a belief fails to be internally justified, then it is reflectively lucky. I will then address the converse, which, as we will see, is not true (this is why the problem of explaining whether or not reflective luck is compatible with knowledge is not *just* the problem of explaining whether or not lack of internal justification entails lack of knowledge).

First, let me make a brief digression on the externalism/internalism controversy in epistemology. In one of the seminal papers that helped inform the debate, Alvin I. Goldman (1979) famously argued that the property of being the output of a reliable belief-forming process is what is at the core of the notion of epistemic justification. In my opinion, part of the controversy arose by and large because Goldman (and others) proposed the reliability requirement (and other externalist constraints) as necessary for justification (and derivatively for knowledge) rather than just simply for knowledge (which seems a more natural move), so that the resulting truth-connected notion of epistemic justification clashed with the reason-giving models of justification traditionally defended by internalists.

Epistemological externalism and internalism can coexist if the former is regarded as an approach to knowledge and the latter as an approach to justification in such a way that (externalistically conceived) knowledge is kept separated from (internalistically conceived) justification.[16] Returning to reflective luck, an externalist following this strategy could consequently think that reflective luck, which arises when true beliefs are not internally justified, pose no problem to her externalist theory of *knowledge* but to internalist theories of *justification*. By way of illustration, consider what Kelly Becker says here:

> I have just distinguished reflective luck as a problem for internalism about justification from veritic luck as an issue for externalist (and internalist) theories of knowledge.

16 The list of epistemologists who implicitly or explicitly hold that justification is not necessary for knowledge is quite extensive. To name a few: Alston (2005), Audi (1998), Becker (2007, 2008), Dretske (1981), Foley (2004), Goldman (1967), Kornblith (2008), Nozick (1981), Plantinga (1993a), Roush (2005). There have also been attempts to bring together internalism and externalism. Notably, Ernest Sosa's virtue epistemology (e.g., 1991) is, among many other things, a remarkable attempt to accommodate the intuitions motivating internalism into a broadly construed externalist theory of knowledge and justification.

> Why the slide from justification to knowledge? The short answer is that, though justification is an important epistemological concept in its own right, once we cast the issue in terms of justification, it threatens to cede too much ground to internalism, where the intuitive concept of justification finds its most natural home. (Becker 2008: 354, fn. 4)

However, externalists cannot get rid of the problem of reflective luck so easily. In general, an archetypical externalist will tend to ascribe knowledge to agents with non-veritically lucky and reliably formed true beliefs, no matter those beliefs are not internally justified and hence reflectively lucky. This would apply to chicken-sexers, plain spotters, clairvoyants and temperature detectors as described in the first set of examples. However, the internalist may oblige the externalist to take up the challenge of reflective luck by modifying the cases in such a way that the force of the intuition that the agents know is diminished.

For example, the internalist could construct cases of reliable chicken-sexers, plain spotters, clairvoyants or temperature detectors who believe at the second-order level that all their past performances have been inaccurate. Alternatively, she could stipulate that the agents are completely mistaken about how they actually form their true beliefs. In both epistemic situations, it is controversial, or at least not so uncontroversial, to claim that the agents know. At that point, the internalist could argue that the reason it is not intuitive to ascribe knowledge to those agents is that in the modified cases the lack of internal justification is even more manifest than in the original cases. Then, the externalist would have no other option but to give an explanation of why the agents fail to know given that their beliefs are non-veritically lucky and reliably formed, that is, given that their beliefs comply with her requirements for knowledge.

§ Is reflective luck compatible with knowledge? It should not be difficult for the externalist to provide such an explanation, for instance, by appealing to epistemic conditions other than reliability. I will not specify them in this chapter, as my aim here is simply to argue, in general, that internalists do not succeed in making a case for the claim that if *S*'s true belief that *p* is not internally justified and hence reflectively lucky, *S* does not know that *p*. That is to say, I aim to argue that reflective luck, at least when its source is the lack of internal justification, is compatible with knowledge.

The condition that in order for *S* to know that *p* *S* must be reflectively aware of the reliability of her cognitive faculties or of the reasons that support her belief that *p* is a too strong requirement. To begin with, there are well-known arguments in the literature that aim to show that an awareness or reflection requirement cannot be necessary for knowledge. In brief: 1) children and animals may acquire knowledge despite they fail to meet the requirement; 2) the relevant second-order beliefs that fix the justificatory status of the relevant first-order beliefs must be themselves internally justified, which leads to an infinite regress.

More importantly, an awareness or reflection condition is too demanding because it rules out the possibility of standard forms of knowledge in mundane situations. This is where the second set of examples becomes relevant. Such a condition would prevent the correct attribution of knowledge to agents who do not recall the evidential source of their true beliefs (Broccoli) or to agents who are not capable of telling by reflection alone how their beliefs have been formed when basic cognitive processes have formed them (Face Recognition). However, the epistemic situations of those agents are too common not to attribute knowledge to them. Not doing so would turn the concept of knowledge into an unnecessarily stringent concept, in that any true belief of an agent who has no reflective access to the belief's justifiers would not amount to knowledge from the outset. Reflective luck, at least when it arises from lack of internal justification, *is* compatible with knowledge.

Nevertheless, although reflective luck can coexist with (externalistically conceived) knowledge, it is worth asking whether the sort of reflective backup that reflectively lucky true beliefs lack is desirable from an epistemic point of view; that is, whether agents with internally justified true beliefs are epistemically better off than agents true beliefs that are not internally justified and hence that are reflectively lucky. For example, is in a better epistemic position a chicken-sexer who is able to tell by reflection alone how she forms her true beliefs about chicks than a chicken-sexer as described in Chicken-Sexer? If so, what epistemically significant property (if any) does the enlightened chicken-sexer have that the naive chicken-sexer lacks?

§ Does reflective luck prevent us from being reliable informants? Pritchard's diagnosis is that naive chicken-sexers are not capable to properly *claim knowledge* of what they believe because they do not meet the relevant *internal* epistemic conditions, to which he adds:

> The ability to properly claim the knowledge that one has is, however, a very desirable epistemic capacity. For whilst it might be useful to us to know that the naïve chicken-sexer is forming safe true beliefs, and thus know that she is a reliable indicator when it comes to the subject matter in question, she herself is not able to perform the role of being a *reliable informant* in this respect, since from her point of view she lacks any reason for thinking that she is forming beliefs in a safe fashion. But the ability to be a reliable informant, to put our knowledge to use in this way, is clearly something of tremendous value to us. (Pritchard 2005: 185)

According to Pritchard, then, reflective luck would undermine our capacity to perform the role of being reliable informants. In what follows, I will show why that is not the case and, in particular, why the following condition is false:

- *S* cannot be in a position to *properly claim knowledge* of what she believes and hence to perform the role of being a *reliable informant* in this respect, unless her beliefs are backed up by reflectively formed second-order beliefs about the reliability of the method by which *S* has formed them or about the reasons in virtue of which *S* holds them.[17]

It is questionable that one cannot perform the role of being a good informant unless one meets internalist standards. As Edward Craig (1990: 62) argues, if we judge the cognitive performance of an agent to be such that it typically yields a high ratio of true beliefs, we, *qua* epistemic community, will judge her to be a good informant quite independently of whether she is herself aware of the correlation with the truth of her beliefs and her cognitive performances.

Turning the objection against internalism, we can argue that, since internal justification does not provide any sort of connection with the truth (one may retain internal justification under massive deception), being internally justified is not necessary to perform the role of being a reliable informant. By contrast, one is not a reliable informant concerning a set of propositions unless externalists requirements (e.g., a reliability condition) are satisfied.

There might still be some wiggle room for internalism. An argument would be that externalism captures what it takes to perform the role of being a good informant only from a third-person perspective: *according to an epistemic group*, *S* is a good informant concerning *p* if *S* has been marked as an approved source of information, and for that the group needs to identify in *S* some property that correlates well with having the truth as to whether *p*. On this third-person perspective, the fact that *S* is aware of such a property is certainly irrelevant to count her as a good informant. But then, in line with Pritchard's quote, the internalist could argue that what is at issue is not whether agents like the naive chicken-sexer are *reliable indicators* of the truth, but whether they can themselves *perform* the role of being reliable informants. To perform such a role, the internalist could argue, they would need reflectively accessible grounds to back up their beliefs. Let us see this in more detail.

As Craig (1990: 65) explains, in some circumstances we may be interested in evaluating whether we ourselves count as good informants. Imagine a situation in which a group needs an informant on

17 This is compatible with the fact that *S* has inductive reasons about the reliability of her method of belief formation. The point is that in order to *properly* claim knowledge of what one believes, one must have reflective reasons to back-up one's first-order beliefs.

some issue and no external evidence is available about who is good enough in that respect. At that point, one member of the group will have to claim knowledge, which may lead her to a self-directed inquiry with the aim of certifying herself as a good informant. This kind of situation is where the internalist standards find their space. In particular, the internalist would claim that the group member would not be capable of performing the role of being a reliable informant unless she is able to internally justify the beliefs that she intends to share.

However, here again internalists requirements are too demanding. In order for the group member to perform such a role, she only needs, besides reliability, confident beliefs, and to have confident beliefs she does not need reflectively accessible grounds for them. Certainly, if she believes at the second-order level that all her past cognitive performances have been inaccurate, she will not be very confident about her actual beliefs. However, it does not follow from that that she needs reflectively accessible grounds for her beliefs in order to perform the role of reliable informant in a confident fashion. To illustrate, chicken-sexers, plain spotters or face recognizers who have no second-order beliefs about how they form their first-order thoughts might confidently and reliably acquire true beliefs (and successfully communicate them) in the absence of awareness of the factors that ground that confidence and reliability. In fact, reflection might be, on occasion, an obstacle to achieving confidence and to the reliable deployment of cognitive performance. The conclusion to be drawn is that reflective luck, when it arises from lack of internal justification, does not undermine neither the capacity to be a reliable informant, nor the possibility of acting as such.

§ Does reflective luck undermine belief ownership? Daniel Breyer (2010), building on previous work (Breyer & Greco 2008) and in response to an objection raised by Bernecker (2008), gives an alternative answer to the question of what epistemically significant property do agents with reflectively lucky beliefs lack. In a nutshell, according to Breyer, reflective luck undermines *belief ownership*. He motivates this idea by focusing on the sort of cases of reflective luck included in set 1. The problem that Breyer sees with these cases is that the relevant beliefs cannot be *attributed* to the agents *qua* epistemic agents. Their beliefs are theirs, of course, in a weak sense of attribution, but not in a strong sense. To exemplify, Breyer (2010: 139) makes the following claim about CLAIRVOYANT: "Norman is a kind of belief machine: what is strange about Norman is not simply that his belief is an accident from his perspective; it's that it's hard to see how we could *attribute* the belief to *Norman*". To make the point more understandable, he draws a useful analogy: "a belief that is not owned in this strong sense (...) lies outside the bounds of cognitive agency in much

the same way that a blink lies outside the bounds of moral agency" (Breyer 2010: 140).

Internalists seem to be in a good position to account for belief ownership. It would suffice for them to put forward some condition of the following type:

- *S*'s belief is attributable to *S* as *S*'s own, only if *S* has reflectively accessible grounds for her belief.

Of course, externalists cannot uphold conditions of this sort. However, this does not mean that they cannot account for belief ownership. As Breyer argues, there are several ways in which externalists can account for it without requiring reflectively accessible grounds. By way of illustration, Breyer's model is based on the idea that by taking responsibility for a certain mechanism (whether moral or cognitive), one thereby 'owns' it and the actions that issue from it:

> In taking epistemic responsibility, one would not need a reflective perspective on the sources of one's beliefs; rather, one need only (i) recognize oneself as the source of one's beliefs, such as perceptual beliefs, (ii) accept that one is fairly credited with having certain beliefs, and (iii) base one's beliefs about oneself on appropriate grounds. (Breyer 2010: 143)[18]

There are more ways of understanding belief ownership that are compatible with epistemological externalism. But we do not need to go into further detail here,[19] as my aim in this section is simply to elucidate in which sense and to which extent reflective luck is epistemically harmful. As regards that aim, we can conclude that reflective luck, interpreted as a phenomenon that undermines belief ownership, is epistemically problematic to certain extent, but both externalists and internalists are in principle in a position to deal with it.

§ *The ineliminability of reflective luck.* Thus far, the problem of reflective luck has been regarded as a problem mainly for externalism. This is because the discussion has focused on cases of agents whose beliefs are not internally justified (cases of sets 1 and 2). In the remaining of this section, I will focus on cases of massive deception and simulated realities such as NEO, brain-in-a-vat scenarios or Cartesian demon worlds; in sum, on the type of cases of set 3. Skeptical scenarios are especially relevant because they show that a belief can be both reflectively lucky *and* internally justified. This point is uncontroversial, as it is a widely spread intuition that agents in skeptical scenarios might be no less justified in their beliefs than us, inhabitants of the actual

18 See also Breyer & Greco (2008).
19 I will devote extensive discussion to belief ownership in chapter 7.

world.[20] Therefore, the problem of reflective luck cannot be just a problem for externalism.

Pritchard develops the idea in a convincing way. He contends that the ascription of internal justification to an agent tends to proceed by implicitly bracketing skeptical possibilities. For example, the claim that the enlightened chicken-sexer is internally justified in her beliefs and hence that he has knowledge can only be held true under the assumption that no skeptical hypothesis obtains. However, Pritchard claims, the enlightened chicken-sexer's reflectively accessible grounds for her beliefs do not prevent them from being reflectively lucky. More specifically, the reasons that the enlightened chicken-sexer has for believing that her first-order beliefs have a positive epistemic status do not defeat the reasons to believe that she is in an skeptical scenario. As a consequence of this, skeptical possible worlds count as *close* in the particular reflective world-ordering needed to establish whether a belief is reflectively lucky. Since in skeptical worlds the enlightened chicken-sexer fails to believe truths, her beliefs in the actual world are reflectively lucky.

Crucially, the connection between reflective luck and skepticism makes the reach of the former as broad as the reach of the latter: to eliminate reflective luck we are obliged to offer reflective grounds that can rule out the possibility of being in an skeptical scenario.

There are several ways to respond to this challenge. Breyer (2010), for instance, thinks that it is illegitimate to make such a demand, for it rests on the assumption that we can be *answerable* for our beliefs *in an ultimate sense*. More specifically, he thinks that it constrains epistemologists to account for "the ultimate reliability of the sources of our beliefs in such a way that they can thereby eliminate the possibility that, from our subjective perspective, our beliefs seems true, even though, from an objective perspective they are false" (Breyer 2010: 149). But to account for that, Breyer argues, ultimate awareness is required, i.e., a sort of God's eye perspective on the ultimate reliability of our sources, which for obvious reasons is a highly implausible requirement.

20 In fact, this is the premise of a much-discussed internalist argument against reliabilist theories of justification. The argument, known as the *New Evil Demon problem*, aims to show that reliability is not a necessary condition for epistemic justification. In evil demon worlds, internalists argue, agents may be as justified in their beliefs as agents in non-skeptical worlds. Take two agents, *S* and *S**, who are identical in every respect, physically and mentally, as Sosa puts it, "not only in respect of mental episodes, but also in respect of deeply lodged dispositions to adduce reasons, etc." (Sosa 2003: 150). Obviously, the reflectively accessible grounds for their beliefs are the same so that *S* and *S** are equally justified (or so is claimed by internalists). However, *S* lives in the actual world and *S** in an evil demon world, which means that, since in the skeptical scenario *S**'s beliefs are not acquired via reliable cognitive processes, reliability cannot be necessary for justification. See Lehrer & Cohen (1983) for the original formulation of the problem.

Pritchard approaches the problem from a different perspective. He first recognizes the ineliminability of reflective luck by pointing out that our practice of offering grounds tends to presuppose the falsity of skeptical hypotheses, which creates the illusion that our grounds can suffice to eliminate reflective luck. He then provides a Wittgensteinian account of why our practice of offering grounds is pragmatically legitimate (the core idea of this account is that such a practice takes place against a backdrop of claims which are of their nature ungroundedly held). According to Pritchard, this account does not provide an *epistemic* response to the problem of reflective luck, but it does provide a *pragmatic* explanation.[21]

What kind of position should we adopt against the problem of reflective luck when connected with skepticism, i.e., when skeptical hypotheses are not excluded from the world-ordering that is relevant to determine whether a belief is true as a matter of reflective luck? We could well embrace one of the two solutions just sketched. However, both answers assume a certain solution to the skeptical problem and a proper answer to the skeptic is beyond the scope of the present project of offering a diagnosis of the problem of luck in epistemology and a theory of knowledge that takes into account what is essential to the concept as we ordinarily apply it, viz., absent any skeptical doubt. In what follows, then, the discussion will proceed under the assumption that skeptical hypothesis are bracketed. Of course, some will find this a dissatisfying move, but it is a legitimate and fairly common move in epistemology to proceed by putting skepticism aside while tackling other important epistemological problems (such as the problem of defining knowledge). As regards the problem of belief ownership posed by reflective luck, we will certainly have to provide a solution, but it will have to wait until chapter 7.

To summarize, section 4.1 has been a thorough analysis of Unger/Pritchard's taxonomy of epistemic luck, which is represented in the following table:

COMPATIBLE WITH KNOWLEDGE	INCOMPATIBLE WITH KNOWLEDGE
Capacity luck	*Veritic intervening luck*
Doxastic luck	*Veritic environmental luck*
Content luck	
Evidential luck	
*Reflective luck**	

* Skeptical hypotheses aside

The outcome of the discussion is, first, that there are two sub-varieties of veritic luck that are epistemically harmful: intervening luck and environmental luck; second, reflective luck is also problematic because

21 See Pritchard (2012a) for further discussion.

it challenges belief ownership, but it is compatible with knowledge unless, third, it is connected with skepticism, which becomes then as threatening as skepticism itself. The first and second problem will be addressed in the next chapters *pace* the third problem, which will be left unsolved for further consideration.

4.2 EPISTEMIC RISK

In part I, I argued that the notion of luck can be explained in terms of two senses of risk: an event-focused sense and an agent-focused sense. We saw that a modal rather than a probabilistic interpretation of the event-focused sense of risk is the one that best fits the notion of luck. I offered the following definition of this sense of risk:

- OBJECTIVE MODAL RISK: *E* is at risk of occurring at *t* if and only if *E* would occur at *t* in a large enough proportion of close possible worlds.

I argued that the threshold beyond which the proportion of close possible worlds is large enough to consider an event risky is a function of the significance that the event has for the agent. For example, although you would not die in most close possible worlds if you played Russian roulette with one bullet in the chamber of a revolver with a 6-shot capacity (the probability of being shot is approximately 16%), we would still intuitively say that your death is at risk of occurring and hence that playing Russian roulette is a risky activity: your life is too significant for you. In addition, I gave the following definition of what is for an agent to be at risk with respect to an event:

- AGENTIAL RISK: *S* is at risk with respect to *E* if and only if (i) *S* has an interest *N*, (ii) if *E* were to occur, it would have some objectively positive or negative effect on *N* and (iii) *S* lacks control over *E*.[22]

AGENTIAL RISK predicts, for example, that you are at risk of dying when playing Russian roulette because you lack control over the outcome of the game. I used both senses of risk to define and distinguish two different applications of the concept to which the terms 'luck' and 'fortune' refer in ordinary discourse (as ordinarily used, there is no much difference between both terms): luck/fortune-I (LF-I) and luck/fortune-II (LF-II), which, for simplicity, I respectively called 'luck' and 'fortune'.

In what follows, I will make a distinction between two senses of *epistemic risk* that is analogous to the event-focused/agent-focused

22 As I explained, AGENTIAL RISK differs from our ordinary notion of risk in that it regards as risky non-controlled events that affect *positively* one's interests, while our ordinary notion of risk is associated only with negative effects. Still, events beyond one's control, even those that carry positive effects, are events on which one cannot count to take further action, which puts one, in a sense, in a risky position.

risk distinction. Epistemic risk has a belief-focused sense (it is a belief that is subject to epistemic risk) and an agent-focused sense (it is the agent and not merely her beliefs who is subject to epistemic risk). These two senses of epistemic risk will allow us to distinguish two general ways in which a true belief might be epistemically lucky, two ways that I will call (in keeping with the technical distinction between luck and fortune mentioned above) *epistemic luck* and *epistemic fortune*. As we will see, only epistemic luck is taken into account (and only partially) in Unger/Pritchard's taxonomy. The distinction will show that the problem of luck in epistemology is more pervasive than it was initially thought.

4.2.1 *Belief-focused Risk*

In the beginning of the chapter, I listed factors that enable or cause the acquisition of knowledge of a proposition *p* by an agent *S:* (1) that *S exists;* (2) that *S* possesses the relevant *cognitive abilities*; that (3) *S forms* the belief that *p*; (4) that there is a *truth-maker* for *p* (or that *p* is true); (5) that *S* has certain *evidence* in favor of the belief that *p* and (6) that *S comes to believe that p truly.* Factors 1-6 might be at risk of not occurring and all of them give rise to different varieties of epistemic risk. However, although all these factors are necessary for knowledge, we have seen that the fact that most of them could easily have not obtained does not necessarily undermine knowledge.

4.2.1.1 *Risk of Error*

Underlying VERITIC LUCK$_{Pritchard}$ there is the idea that an agent does not know that *p* if her true belief that *p* was at (high) risk of being false. However, this idea needs qualification. An epistemic analogue of OBJECTIVE MODAL RISK when applied to factor 6 would be roughly the following: suppose that *S* comes to believe that *p* truly; *S*'s coming to believe that *p* truly is risky just in case *S* would not come to believe that *p* truly in a large enough proportion of close possible worlds. How large should that proportion be? It is not easy to give a precise answer, although we would not say that a belief is epistemically risky (at least in a significant sense) if the proportion of close possible worlds in which it would be false is a small one. In order for a belief to be at significant risk of being false, it should be false in at least more than half the close possible worlds. Accordingly, we can formulate an epistemic analogue of OBJECTIVE MODAL RISK in the following way:

- UNRESTRICTED BELIEF-FOCUSED RISK I: *S*'s belief that *p* is risky$_1$ if and only if *p* is false in at least half the close possible worlds in which *S* comes to believe that *p*.

A belief can be epistemically risky$_1$ and be knowledge. To see this, consider again GRANDMOTHER. Intuitively, the grandmother knows that her grandson is alive when she looks at him, although the grandson could easily have died and she could easily have formed the false belief that he was alive (because others would have told her so). That is to say, she knows that her grandson is alive even though in more than half the close possible worlds in which she believes that her grandson is alive, her grandson is dead.

As we saw, VERITIC LUCK$_{Pritchard}$ correctly predicts that the grandmother's belief is *not* lucky. VERITIC LUCK$_{Pritchard}$ says that *S*'s true belief that *p* is luckily true just in case (1) *p* is false in most close possible worlds in which *S* continues to believe that *p* and (2) in which the relevant initial conditions for the formation of that belief are the *same* as in the actual world. Close possible worlds in which the grandmother believes falsely that her grandson is alive (most close possible worlds) are worlds in which she forms her belief by testimony, whereas in the actual world she uses a visual method (hence (2) does not hold).

VERITIC LUCK$_{Pritchard}$ takes then into account a particular type of knowledge-undermining risk, which we can state in the following way:

- BELIEF-FOCUSED RISK I: *S* does not know that *p* in the actual world if in at least half the close possible worlds in which *S* continues to believe *p* and the relevant initial conditions for *S*'s belief that *p* are the same as in the actual world, *p* is false.

BELIEF-FOCUSED RISK I explains why in case F, a case of veritic luck, Adam does not know that there is a snake in the tree: because in most close possible worlds in which he continues to believe that there is snake in the tree and the relevant initial conditions for that belief (looking at the snake-looking branch) are the same as in the actual world, the snake does not climb the tree and hence his belief is false. Another example explained by BELIEF-FOCUSED RISK I is the first Gettier-style that appeared in the literature:

> STOPPED CLOCK
>
> There is the man who looks at a clock which is not going, though he thinks it is, and who happens to look at it at the moment when it is right; this man acquires a true belief as to the time of day, but cannot be said to have knowledge. (Russell 1948: 170-171)

That man's belief is not knowledge because in close possible worlds in which the man forms the belief in the same way (by looking at the clock), he does it at a slightly different time and, consequently, the man forms a false belief in the same proposition. In general, BELIEF-FOCUSED RISK I explains why many Gettier-style cases are not cases of knowledge.[23]

23 For an elucidation of why I say *many* and not *all* Gettier-style cases, see section 4.3.

4.2.1.2 *Risk of Close Error*

There are other types of epistemic risk and, in turn, other senses of veritic luck not covered by Pritchard's definition. David Manley (2007) identifies another way in which *S* could have failed at the doxastic level: *S* could have believed closely related false propositions.[24] He gives the following example:

> Bird
>
> I have a true demonstrative thought, one that I express by saying 'That is a lark'. My thought is true: *that* is a lark. However, there are many lark-imitating imposters nearby. (Adapted from Manley 2007: 404)

Another example:

> Red Barn
>
> Carl enters an area populated by yellow fake barns and looks at the only real barn, which is red. He forms the belief that he's driving by a red barn. (Adapted from Pryor 2004: 71)

The relevant sense of epistemic risk in Bird and Red Barn is not the one defined by Belief-focused Risk I. According to Manley, his belief in the proposition that *that* is a lark, could *not* have been false, because he assumes "that the content of a demonstrative thought is a proposition involving some object *a* whose truth value with respect to any possible world depends solely on how things stand with *a* at that world" (Manley 2007: 404). That is, since it is true that *that* is a lark in the actual world, the proposition that *that* is a lark is true in all possible worlds in which *that* bird exists and at least not false in worlds where it does not. Perhaps Manley would not come to believe that proposition in any close possible world, but that is sufficient to conclude that his actual belief is not risky in the sense of Belief-focused Risk I: in close possible worlds he does not come to believe the *same proposition* falsely.

In Red Barn, Carl does not know that he is driving by a red barn when he forms the belief that he is driving by a red barn. At best, he knows that he is driving by a red object, but not by a red *barn*. The important point is that if he came to believe that he is driving by a red barn in close possible worlds, that belief would be true (he would look at the only red barn in the area). In most close possible worlds, however, he *does* form a false belief, but not in the same proposition: most close possible worlds are scenarios in which he drives by yellow fake barns, looks at them, and forms false beliefs that he is driving

24 See Hiller & Neta (2007): they argue that Pritchard's definition of epistemic luck fails to account for several cases that can be explained in terms of this kind of belief-focused risk.

by yellow barns. In other words, in most close possible worlds Carl believes closely related false propositions. Therefore, Carl's actual belief cannot be risky in the sense specified by BELIEF-FOCUSED RISK I. In general, the sort of belief-focused risk involved in BIRD and RED BARN is the risk of coming to believe *closely related false propositions*.

It is not easy to specify the notion of closely related proposition. The best we can do is to get an intuitive grasp with a couple of examples. Plato provides an excellent one:

> *Socrates*: When at such a stage in his progress a person in writing 'Theaetetus' thinks he ought to write, and actually does write, TH and E, and again in trying to write 'Theodorus' thinks he ought to write, and does write, T and E, shall we say that he knows the first syllable of your names? [Socrates speaks to Theodorus and Theaetetus]
>
> *Theaetetus*: No, we just now agreed that a person in such a condition has not yet gained knowledge.[25]

Believing that 'Theodorus' starts with T and E is a false proposition closely related to the true proposition that 'Theaetetus' starts with TH and E. Similar thoughts apply to propositions about colors. Suppose that someone believes that certain object is painted red by looking at it. Suppose that the same person, in similar circumstances, believes that a very similar object painted with a very similar shade of red thinks that the object is yellow. Does this person know that the first object is painted red? There is good reason to doubt that. In the same way, we do not typically attribute knowledge to an agent who believes that there is an F (e.g., a (red) barn) when F is before her eyes if she would also believe that there is an F when a relevantly similar object that is not an F (e.g., a (yellow) fake barn) is before her eyes. All this shows is that our cognitive abilities and methods of knowing are relativized to *fields* of closely related propositions, to the extent that if one knows a proposition in a field via some method M, one must not believe at a minimum the negations of several propositions in that field via M. If one did, one's belief in that proposition would not be knowledge.[26]

BIRD and RED BARN are cases of veritic luck but that is not taken into account by VERITIC LUCK$_{Pritchard}$, which defines veritic luck only in terms BELIEF-FOCUSED RISK I. The kind of epistemic risk in BIRD and RED BARN is rather the following:

25 Theaetetus (207e-208a), from Fowler (1921).

26 The idea that our cognitive abilities and methods of knowing are relativized to fields of propositions is held by anyone who holds a global reliability condition for knowledge, the condition that if one knows that p, one's belief that p is the output of a belief-forming process or method that is reliable not only with respect to p but to a wider range of propositions. Most reliabilists and virtue epistemologists share this basic idea. See Goldman (1986: ch. 3), Greco (2000: 216) and Sosa (1991: ch. 16) for relevant discussion.

- BELIEF-FOCUSED RISK II: *S* does not know that *p* in the actual world if in at least half the close possible worlds *S* would form beliefs in false propositions closely related to *p* (were the relevant initial conditions for those beliefs to be the same as the initial conditions for *S*'s belief that *p* in the actual world).

Finally, let me stress the importance of the initial conditions clause of BELIEF-FOCUSED RISK II. Suppose that I come to believe by some mental method *M* that 2+2 = 4. *M* is globally reliable and has never led me to believe that 2+2 = 5 or that 2+3 = 4. On this occasion, I could easily have been administered a drug that would have made me believe that 2+2 = 5 or that 2+3 = 4. This fact, however, does not prevent me from knowing that that 2+2 = 4 via *M*, as in my present situation I have not been yet administered any drug. In other words, the modal fragility of my epistemic position does not prevent me from knowing that 2+2=4. Of course, in most close possible worlds I would believe some closely related false proposition (that 2+2 = 5 or that 2+3 = 4), but the relevant initial conditions for my beliefs in those worlds would not be the same as the relevant initial conditions for my belief in the actual world. We can accordingly distinguish another belief-focused sense of risk that is compatible with knowledge possession:

- UNRESTRICTED BELIEF-FOCUSED RISK II*:* *S*'s belief that *p* is risky$_2$ if and only if in at least half the close possible worlds *S* would form beliefs in false propositions closely related to *p.*

As we will see in chapter 9, this kind of distinctions between epistemically innocuous senses of belief-focused risk and the knowledge-undermining corresponding senses will be relevant in the defense of the safety condition for knowledge.

4.2.1.3 *Further Risks*

Manley (2007) distinguishes some other types of epistemic risk. He uses, to that aim, cases of hallucination like the following:

> HALLUCINATING BIRDS
>
> Suppose that (unwittingly) I often hallucinate lark calls. In between some hallucinations, I hear a real lark and form the demonstrative thought I express by 'That bird is a lark'. (Manley 2007: 404)

Depending on one's view on hallucinations, the kind of epistemic risk at issue here will be differently conceived. For example, if hallucinatory experiences have no content at all, as some philosophers maintain, then the sort of epistemic risk in HALLUCINATING BIRDS is the risk of forming a belief with no content. By contrast, if hallucinatory experiences have gappy content of the form '___ is a lark', then the relevant epistemic risk is the risk of forming a belief with gappy content. In sum, another type of epistemic risk is the risk of forming a belief with gappy content or with no content at all.

Manley distinguishes a further type of epistemic risk: the risk of forming a belief with paradoxical content. He motivates the distinction with the two following examples:

> PARADOX I
>
> Suppose Mary believes that what Susan just wrote in her diary is false. Aside from this belief, all of Mary's beliefs about Susan are true. Suppose Susan actually wrote something false (e.g., '2+2 = 5'), had no inclination to write anything true, but nearly wrote something paradoxical such as 'Everything Mary believes about me is true'. Then Mary's belief is true, and she is not in danger of having a false belief, but she is in danger of epistemic failure. (Adapted from Manley 2007: 405)
>
> PARADOX II
>
> *X* is an inhabitant of a world very sparsely populated in which people spend much of their time in deep sleep. *X* forms the belief that someone is thinking something untrue right now. In fact, someone else, *Y*, is thinking something untrue right now. (Adapted from Manley 2007: 405-6)

Susan could easily have written in her diary 'Everything Mary believes about me is true' and hence Mary could easily have formed a belief in the paradoxical proposition that what Susan just wrote in her diary ('Everything Mary believes about me is true') is false. On the other hand, *Y* could easily have been asleep and therefore *X*'s belief that someone is thinking something untrue could easily have been paradoxical (in that hypothetical case, *X* would have been the only person thinking something). As Manley correctly points out, in all these cases there is risk of epistemic failure.

However, it is not clear that, as the cases are described, the relevant epistemic risk is incompatible with knowledge. Consider HALLUCINATING BIRDS. As the case is described, Manley seems to be suffering hallucinations all the time and, from time to time, he hears a real lark. In that case, there is obviously not knowledge. But the following reading is compatible with the way the case is described: suppose that Manley comes to believe that that bird is a lark when he hears a real lark because the drug that was unwittingly making him hallucinate larks has ceased to have effect on him thanks to the ingestion of some pill that thwarts the effects of the drug *momentarily*. Manley's epistemic position is modally fragile, of course, but, arguably, he can know that that bird is a lark because he forms his belief via his normally reliable and properly functioning perceptual abilities, whereas in close possible worlds he would form the relevant beliefs in a different way, i.e., under the effects of the drug.

Analogously, a reading of PARADOX II that would deliver a verdict of knowledge would be the following: suppose that *X* forms her true belief that someone is thinking something untrue right now on the basis of seeing *Y* write '2+2=5' on a blackboard. Suppose then that there is an easily possible counterfactual scenario in which *Y* faints and consequently entertains no thought at all. In that scenario, let us stipulate, although *X* would not see that *Y* is writing '2+2=5' on the blackboard, she would still form the belief that someone is thinking something untrue in some other way and nobody, except *X*, would be thinking something. Under this description of the case, which seems compatible with how PARADOX II is described, it is not clear that *X* does not know that someone is thinking something untrue right now when seeing *Y* write '2+2=5' on a blackboard in the actual world. Similar considerations apply to PARADOX I.

The point is that, as before, we can distinguish another unrestricted sense of epistemic risk that seems to be compatible with knowledge. As Manley (2007: 406) suggests, we can define a counterpart relation on thoughts or beliefs, broader than identity, such that:

- UNRESTRICTED BELIEF-FOCUSED RISK III*:* *S*'s belief that *p* is risky$_3$ if and only if in at least half the close possible worlds *S* forms a counterpart belief in a proposition with gappy content, paradoxical content or with no content at all.

Cases like HALLUCINATING BIRDS, PARADOX I and PARADOX II are not cases of knowledge if the way they are described makes explicit that the relevant initial conditions for the relevant counterpart failed beliefs are the same as the initial conditions for the target actual beliefs. The third sense of knowledge-undermining risk can be then stated as follows:

- BELIEF-FOCUSED RISK III: *S* does not know that *p* in the actual world if in at least half the close possible worlds *S* would form a counterpart belief in a proposition with gappy content, paradoxical content or with no content at all and if the relevant initial conditions for that belief are the same as the initial conditions for her belief that *p* in the actual world.

Again, we will appreciate the value of these distinctions in chapter 9, when the safety condition for knowledge will be defended from a recent objection.

4.2.2 *Agent-focused Epistemic Risk*

The technical distinction between luck and fortune offered in part I was motivated by the important difference existing between lucky/

fortunate events such as winning a fair lottery and lucky/fortunate events such as winning a lottery rigged in one's favor: in the former case one could easily have lost the lottery, while in the latter it would be almost impossible for one to lose it. Still, both events are considered lucky or fortunate insofar as one does not control the lottery process and its outcomes. Our ordinary usage of the terms 'luck' and 'fortune' overlooks this difference and I decided to call (technically) the first sense of 'luckiness' or 'fortunateness' *luck* and the second sense *fortune*. Here are the definitions that I gave:

- LUCK: E is lucky for S if only if: (i) S has an interest N and E has some objectively positive or negative effect on N (in the sense that E is good or bad for S), (ii) E occurs in the actual world but would not occur in a large enough proportion of close possible worlds where the relevant initial conditions for E are the same as in the actual world and (iii) S lacks control over E.

- FORTUNE: E is fortunate for S if only if: (i) S has an interest N and E has some objectively positive or negative effect on N (in the sense that E is good or bad for S), (ii) E occurs in the actual world and would occur in a large enough proportion of close possible worlds where the relevant initial conditions for E are the same as in the actual world and (iii) S lacks control over E.[27]

Pritchard's account of luck takes the degree of risk of nonoccurrence that an event had before occurring to be what is essential to the notion of luck as ordinarily understood. However, cases such as winning a rigged lottery prove this account, if not wrong, at least incomplete. The kind of 'luckiness' or 'fortunateness' that we experience when wen win both a fair and a rigged lottery is fundamentally explained in terms of the agent's lack of control over the lottery process and over its significant outcomes, while the difference in 'luckiness' or 'fortunateness' between winning a fair lottery and winning a rigged lottery is explained by the high risk that we have of losing in the former case and the absence of such a risk in the latter. This is, I argue, the correct account of luck (and fortune).

Then why should we only appeal, as Pritchard does, to the modal fragility of an agent's true belief to account for epistemic luck when the lack of modal robustness of an event (in general) is not what is essential to its being lucky? If what is at the core of the concept to which the terms 'luck' and 'fortune' refer in ordinary discourse is fundamentally the fact that the relevant significant event is beyond the agent's control, it is reasonable to think that there is an epistemic

27 We must always keep in mind that LUCK and FORTUNE do not pick out the ordinary notions of luck and fortune, which are the same thing from the layman's point of view. LUCK and FORTUNE are rather Carnapian explications of the concept to which the terms 'luck' and 'fortune' refer when ordinarily used, i.e., luck and fortune are similar, exact, fruitful and simple *explicanda* of that concept.

analogue of the lack of control condition that can account for the notion of epistemic luck. To put it differently, why cannot we claim that epistemic luck arises when an agent lacks some form of epistemic control and that the modal fragility of her beliefs is only *one way* in which they could be epistemically lucky?

There is an interesting sense of epistemic risk that is not exhausted by the different types of belief-focused risk distinguished above and that concerns several attributes of the agent and not merely her beliefs and, in this sense, it is an *agent-focused* form of epistemic risk. My contention is that an agent is at epistemic risk when she lacks *epistemic control*. How should we understand the notion of epistemic control? There are several ways to unpack it. For example, one may find acceptable to claim (from a pronounced internalist perspective) that epistemic control consists in deciding reflectively, voluntarily and directly what to believe or disbelieve. Curiously, Pritchard has only this kind of perspective in mind when he quickly rejects the plausibility of an epistemic lack of control condition by appealing to the fact that one's most basic perceptual beliefs are not "within one's immediate control" (Pritchard 2005: 127).

However, from an externalist perspective epistemic control might be seen as a phenomenon that mostly proceeds in an automatic, unreflective, involuntary way and that, only occasionally, takes place in an indirect reflective manner. This sense of epistemic control is certainly compatible with our intuitive picture of how we acquire basic perceptual beliefs and hence epistemic control so conceived would not have the consequence of regarding basic perceptual beliefs as inherently lucky, as Pritchard thinks it follows from the inclusion of an epistemic lack of control condition in the definition of epistemic luck.

I do not intend to develop an externalist notion of epistemic control here (a complete picture of epistemic control will only be achieved at the end of chapter 9). For the moment, I just want to motivate the notion and show that there are several possible ways of unpacking it and that none of them seems to be neutral as regards what is the correct theory of knowledge. In my view, the best framework to elaborate the concept is provided by virtue epistemology (in particular, by its reliabilist branch) and by modal epistemology. However, it is up to epistemologists to unfold the notion of epistemic control within the theoretical framework that they find most plausible (classical internalists included).

§ *An objection considered*. Let me consider a general objection to the present approach to epistemic luck. The objection casts doubt on the benefits of including the notion of epistemic control in the definition of epistemic luck. It could be argued that the notion needs theoretical elaboration to be properly understood and that its theoretical complexity would transform the notion of epistemic luck into a too heavily theoretically-loaded concept. It should be preferable, the objection continues, to count with a more simple definition of epistemic luck that can be applied without many theoretical assumptions. VERITIC LUCK$_{Pritchard}$, for example, is a good candidate, as it does not presuppose any more theoretical background than minimal assumptions about closeness and possible worlds.

Let me argue just the opposite. The theoretical richness of the notion of control makes the inclusion of an epistemic analogue of the lack of control condition into the definition of epistemic luck something desirable. For one thing, it allows us to appreciate the deep reach of the phenomenon by allowing us to distinguish forms of knowledge-undermining luck not distinguished in Pritchard's account. For another, it allows us to see a significant variety of theories of knowledge as different approaches to what can be described as *anti-luck epistemology*. Pritchard's modal account of epistemic luck and particularly his definition of veritic luck is very restrictive in this regard.

By way of illustration, according to Pritchard (2012b: 249), virtue epistemology does not have as aim, at least not as one of its main aims, the exclusion of epistemic luck. For Pritchard, only purely modal accounts of knowledge such as sensitivity-based and especially safety-based accounts of knowledge can be properly said to be anti-luck epistemologies. However, this is an unfair interpretation of many theories of knowledge that in their attempts to give a solution to the Gettier problem (viz., the problem of explaining why Gettier subjects do not know) put forward valuable epistemic conditions such as (global) reliabilist, virtue-theoretic or even internalist conditions.

The project of giving a solution to the Gettier problem is to a large extent the project of supplementing one's definition of knowledge with an anti-luck (anti-Gettier) condition. Because of this, my position is just the opposite to the simplistic position on epistemic luck that the proponent of the objection suggests to adopt: it is fully desirable to include the notion of epistemic control in the definition of epistemic luck because, given that the notion of epistemic control can be developed within very different theoretical frameworks, many theories of knowledge can be interpreted as giving a response to the problem of luck in epistemology, which unifies, I think, the historical development of the discipline since Gettier's challenge to the tripartite definition of knowledge.[28]

28 Pritchard (2012b) makes a different reconstruction of the debate. He thinks that what he calls *the analytical project* —"the project of offering an adequate definition of knowledge, one that was informative, non-circular, and which could suitably accommodate our salient epistemological intuitions" (Pritchard 2012b: 247)— is motivated by two 'master' intuitions: the so-called *anti-luck intuition*, "the intuition that that when one knows, one's cognitive success (that is, one's believing truly) is not a matter of luck" (*ibid.*), and the *ability intuition*, "the intuition that knowledge requires cognitive ability, in the sense that when one knows, one's cognitive success should be the product of one's cognitive ability" (*ibid.*) (Pritchard's account will be thoroughly analyzed in section 8.2). The notion of epistemic control, as we will see, accommodates both intu-

But more importantly, as we will see in the next section, if we do not include an epistemic analogue of the lack of control condition in our definition of veritic luck, we will not be able to explain some Gettier-style cases (cases which are traditionally considered paradigmatic examples of epistemic luck) and many other cases as instances of veritic luck. Pritchard's definition of veritic luck gets the wrong results in this respect. The way to overcome this obstacle is, first, as we saw in the previous section, to take into account all the senses in which a belief might be epistemically risky and, second, to supplement the definition of veritic luck with an epistemic analogue of the lack of control condition. Only in this way we will be able to appreciate the overwhelming reach of the problem of luck in epistemology.

4.3 EPISTEMIC LUCK AND EPISTEMIC FORTUNE

Let me briefly recap. It is not the same to win a lottery fairly as to win a lottery that has been rigged, unbeknownst to one, in one's favor. I have explained the divergence of intuitions by claiming (technically) that in the former case one is lucky to win while in the latter one is simply fortunate. In both cases we would say that there is 'luckiness' or 'fortunateness' involved because one lacks control over the lottery process (when fair and when manipulated). Now, if we are to introduce an epistemic analogue of the lack of control condition in our definition of veritic luck, we should expect that the set of cases that intuitively instantiate knowledge-undermining luck splits into two subsets: the subset of cases that are structurally equivalent to cases of luck and the subset of cases that are structurally equivalent to cases of fortune. We have already seen a lot of examples of *veritic luck*. However, are there cases of *veritic fortune*? If there are, then our account is on the right track.

Consider the following modification of case F, the example of veritic luck given above:

> STILL SNAKE
>
> Adam sees a snake-looking thing in the tree. The thing Adam is looking at is in reality a snake-looking branch. On the unseen side of the trunk, there is a snake, which has always lived in that tree and has never come down. Adam's belief is true, but, intuitively, he does not know that there is snake in the tree.

Adam's belief is true by luck. However, this is not something that follows from Pritchard's definition of veritic luck. VERITIC LUCK$_{Pritchard}$ says that a belief is true by luck just in case it would be false in most close possible worlds in which the relevant initial conditions for the

itions and, as a consequence, so does my account of epistemic luck. In my view, the problem of luck in epistemology is much more extensive than is commonly thought.

belief are the same as in the actual world. In STILL SNAKE, there is no close world (at least not a wide class of them) in which Adam believes that there is a snake in the tree and his belief is false.

Perhaps the problem is that VERITIC LUCK$_{Pritchard}$ fails to rule that STILL SNAKE is a case of veritic luck because it takes into consideration only one sense of belief-focused risk (namely BELIEF-FOCUSED RISK I). Certainly, VERITIC LUCK$_{Pritchard}$ does not cover cases in which an agent's belief is veritically lucky because of risky in the other senses of belief-focused risk, i.e., in the sense that the agent could easily have formed a false belief in a closely related proposition or a belief with gappy content, paradoxical content or with no content at all. Unfortunately, STILL SNAKE is not such a case. As the example is described, nothing indicates that there is knowledge-undermining belief-focused risk involved, not even epistemically harmless belief-focused risk (i.e., UNRESTRICTED BELIEF-FOCUSED RISK I, II or III).[29]

In sum, we have found a *Gettier-style case* that involves knowledge-undermining luck but no belief-focused risk. How could we account for this distinctive type of epistemic luck? A plausible way is to appeal to the aforementioned notion of agent-focused epistemic risk (i.e., epistemic risk that concerns several attributes of the agent and not merely the modal profile of her beliefs) and to include, accordingly, an epistemic analogue of the lack of control condition in our definition of veritic luck.

Again, I will not elaborate the notion of epistemic control here. My aim, for the moment, is merely to make a distinction between two different epistemic phenomena, which mirrors my previous distinction between luck and fortune. On the one hand, we have several cases in which there is belief-focused risk involved, such as case F, FAKE BARNS, STOPPED CLOCK, BIRD or RED BARN. In all these cases, the agents in question could easily have suffered some form of epistemic failure: in most close possible worlds they would have believed the same proposition falsely or closely related false propositions or propositions with gappy content, paradoxical content or with no content at all. Their beliefs instantiate what I will henceforth call *veritic epistemic luck*. To put it roughly, an agent's true belief is veritically lucky just in case the agent lacks epistemic control and her true belief is epistemically risky in a knowledge-undermining belief-focused sense (i.e., in the sense of BELIEF-FOCUSED RISK I, II or III). More precisely:

29 One could argue that there is BELIEF-FOCUSED RISK II, because Adam could easily have formed the belief that there is a snake in the *visible* part of the tree, i.e., a belief in a closely related *false* proposition. We can exclude this kind of risk from the example by modifying it in a number of ways, e.g., by stipulating that Adam does not possess the concept of visible, or that Adam has ingested a pill that prevents him from forming such a belief, or, better, by stipulating that God would kill Adam if he were to form such a belief (suppose that God does not like the word 'visible'). Similar stipulations could be made about any other belief in any other closely related false proposition.

- Veritic Luck: S's true belief that p is *veritically lucky* if and only if:

    1. in at least half the close possible worlds in which S continues to believe p and the relevant initial conditions for that belief are the same as in the actual world, p is false *or*
    2. in at least half the close possible worlds S would form beliefs in false propositions closely related to p (were the relevant initial conditions for those beliefs to be the same as the initial conditions for S's belief that p in the actual world) *or*
    3. in at least half the close possible worlds S would form a counterpart belief in a proposition with gappy content, paradoxical content or with no content at all (were the relevant initial conditions for that belief to be the same as the initial conditions for S's belief that p in the actual world) *and*
    4. S lacks epistemic control.

As we have seen, Still Snake is not a case of veritic luck because conditions (1), (2) and (3) do not hold. Still, we have the intuition that the case involves some epistemic 'luckiness' or 'fortunateness'. After all, it is a Gettier-style case. The kind of 'luckiness' or 'fortunateness' involved is what I will henceforth call *veritic epistemic fortune.* Roughly, an agent's true belief is veritically fortunate just in case the agent lacks epistemic control and her true belief is *not* epistemically risky in any knowledge-undermining belief-focused sense. More precisely:

- Veritic Fortune: S's true belief that p is *veritically fortunate* if and only if:

    1. in more than half the close possible worlds in which S continues to believe p and the relevant initial conditions for S's belief that p are the same as in the actual world, p is *true and*
    2. in more than half the close possible worlds S would *not* form beliefs in false propositions closely related to p (were the relevant initial conditions for those beliefs to be the same as the initial conditions for S's belief that p in the actual world) *and*
    3. in more than half the close possible worlds S would *not* form a counterpart belief in a proposition with gappy content, paradoxical content or with no content at all (were the relevant initial conditions for that belief to be the same as the initial conditions for S's belief that p in the actual world) *and*

4. *S* lacks epistemic control.[30]

Consider more examples of veritic fortune:

> TEMP
>
> [O]ur agent—let's call him 'Temp'—forms his beliefs about the temperature in his room by consulting a thermometer on the wall. Unbeknownst to Temp, however, the thermometer is broken and is fluctuating randomly within a given range. Nonetheless, Temp never forms a false belief about the temperature by consulting this thermometer since there is a person hidden in the room, next to the thermostat, whose job it is to ensure that whenever Temp consults the thermometer the temperature in the room corresponds to the reading on the thermometer. (Pritchard *et al.* 2010: 48-49)

Temp does not know the temperature of the room, but he could not easily have been mistaken about it. In a sense, he is lucky that an agent modifies the temperature in such a way that makes it correspond to the reading of the broken thermometer whenever Temp consults it. In the absence of belief-focused risk, Temp's beliefs are therefore veritically fortunate.[31]

> MARY
>
> Mary sees someone who looks just like her husband sitting in his favorite chair in the living room and comes to believe that her husband is in the living room. Her husband is in the corner of the living room, unseen by Mary. Mary not only forms the true belief that her husband is in the living room, but imagine as well that her husband's doppelganger knows that she will form such a belief. Further, (. . .) the doppelganger has decided to only appear in the favorite chair when Mary's husband is, unbeknownst to Mary, also in the room. (Brueckner & Buford 2013; adapted from Zagzebski 1996)

30 As we see, VERITIC LUCK and VERITIC FORTUNE do not include, unlike LUCK and FORTUNE, a significance condition. The significance condition says that an event (or state) is lucky or fortunate for *S* only if (i) *S* has an interest *N* and (ii) that event or state has some objectively positive or negative effect on *N*. In the epistemic case, the significance condition cannot remain unsatisfied because the lucky state in question is a doxastic state. Consequently, the fact that a belief is *your* belief makes the belief significant for you, at least minimally. Compare this fact with our difficulty in part I to find an example of non-significant event: the mere fact of considering an event as an example already ascribes a minimal degree of significance to the event.

31 It is important to keep in mind that the hidden agent modifies the temperature of the room and not the values shown on the screen of the thermometer. If the agent manipulated the thermometer instead of the temperature, the intuition that there is no knowledge would not be so strong. In fact, one could argue that the case, so described, would be analogous to a case of testimonial knowledge.

As Brueckner and Buford point out, MARY is not a case of knowledge (Mary mistakes a doppelganger of her husband for her husband, although her belief is in the end true). In addition, the case is a clear case of knowledge-undermining luck. Yet, the kind of luck involved is not that specified by VERITIC LUCK because Mary's situation is such that she could *not* easily have failed from an epistemic point of view: her actual true belief that her husband is in the room could not easily have been false, and she could not easily have formed a false belief in a closely related proposition or a belief with gappy content, paradoxical content or with no content at all.[32] In the absence of belief-focused risk, Mary's true belief is, therefore, veritically fortunate.

4.3.1 *Under-described Cases*

Sometimes, familiar cases are not easy to categorize because they are described in a way that does not make explicit the degree of belief-focused risk involved. Consider a very famous case:

> COINS
>
> Suppose that Smith and Jones have applied for a certain job. And suppose that Smith has strong evidence for the following conjunctive proposition: (d) Jones is the man who will get the job, and Jones has ten coins in his pocket. Smith's evidence for (d) might be that the president of the company assured him that Jones would in the end be selected, and that he, Smith, had counted the coins in Jones's pocket ten minutes ago. Proposition (d) entails: (e) The man who will get the job has ten coins in his pocket. Let us suppose that Smith sees the entailment from (d) to (e), and accepts (e) on the grounds of (d), for which he has strong evidence. (...) But imagine, further, that unknown to Smith, he himself, not Jones, will get the job. And, also, unknown to Smith, he himself has ten coins in his pocket. (Gettier 1963: 122)

This case by Gettier is typically presented as a paradigmatic example of epistemic luck. Yet, we do not know whether it is a case of veritic luck or of veritic fortune. In particular, it is not clear whether Smith's belief could easily have been false. Suppose that the president of the company was completely determined to give the job to Smith but he wanted to play a joke on him, and suppose further that Smith's wife puts, unbeknownst to Smith, ten coins in his pockets every morning. In that case, Smith's belief that the man who will get the job has

32 As before (see fn. 29), one could argue that Mary could easily have formed the false belief that the visible person is her husband. As before, similar modifications to the case can be made to block this sort of possibilities.

ten coins in his pocket could not easily have been false, as in close possible worlds in which Smith forms that belief, the belief is true. The case, so specified, would be thus a case of veritic fortune. But other specifications would make it a case of veritic luck. Therefore, Gettier's example, as it stands, is under-described.

4.3.2 *An Overarching Hypothesis about Knowledge*

The extent to which veritic luck and veritic fortune are epistemically problematic is overwhelming. I take the following hypothesis to be very plausible:

- Hypothesis: nearly all (if not all) cases of true belief that is not knowledge are cases either of veritic luck or of veritic fortune.

The variety of cases that we have analyzed so far supports the hypothesis. This hypothesis, if correct, identifies a common factor shared by nearly all (if not all) counterexamples to the great variety of analysis of knowledge that have appeared in the the literature since the 60's. I do not claim (yet) that all cases of not known true belief are either cases of veritic luck or veritic fortune, because some cases might be difficult to accommodate. Let us consider two type of cases.

From an externalist point of view, it is plausible to think that chicken-sexers, plane spotters, clairvoyants and temperature detectors can have knowledge if they meet the type of externalist requirements for belief ownership that Breyer sketches in his discussion of reflective luck (as well as other requirements for cognitive integration).[33] Consider, however, the following case of true belief which does not seem to be a case of knowledge:

> Misled Chicken-Sexer
>
> A professional chicken-sexer looks at a chick and forms the true belief that it is a male. The chicken-sexer is unaware of the process by which she tells the sex of the chick, but, as always, she is correct in her judgment. In addition, she believes that all her past judgments about the sex of chicks have been incorrect.

The chicken-sexer's true belief is not veritically lucky because she could not easily have been subject of any form of epistemic failure and yet she does not have knowledge. If this is a case of veritic fortune, then the lack of control condition must explain why the chicken-sexer does not know. However, in which sense does the fact that one believes that all one's past judgments have been incorrect makes one lack epistemic control? The question cries out for an answer.[34]

33 See section 7.3.
34 For the curious reader: the answer is in section 7.3.2, fn. 10.

Another type of cases that might be difficult to accommodate are lottery cases. You do not know that your ticket is not the winner despite the high probability you had of not winning. Presumably, given that probability your true belief that your ticket is not a winning ticket could not easily have failed. If this is a case of veritic fortune, then the lack of control condition should explain why your belief is not knowledge. Can the lack of control condition account for the difference between this case and cases of correct inductive inference, which are, presumably, cases of knowledge? Another open question that needs a answer.[35]

4.4 SUMMARY

Let us take stock. I started by presenting Unger/Pritchard's classification of 'epistemic' accidents and the corresponding distinction between harmful and harmless varieties of epistemic luck. I then introduced Pritchard's definition of veritic luck as well as his distinction between intervening and environmental luck. After that, I moved on to discussing the extent to which reflective luck is epistemically problematic. The conclusion was that, if we do not include skeptical hypotheses in the ordering of worlds that is needed to establish whether our beliefs are reflectively lucky, then reflective luck is problematic insofar as it might challenge the ownership of our beliefs. If we include skeptical hypotheses in the relevant world-ordering, then eliminating reflective luck from our theory of knowledge amounts to eliminating the possibility that we are in a massively deceptive skeptical scenario and hence that no ascription of propositional knowledge about the external world is true. If so, we might not be able to eliminate reflective luck, because it is far from clear that we can establish that we are not in an skeptical scenario. In brief, if one thinks that reflective luck is compatible with knowledge, it is because one brackets skeptical hypotheses, something that both internalists and externalists typically do when it comes to define what knowledge is.[36]

I then analyzed several types of epistemic risk and showed that Unger/Pritchard's taxonomy of epistemic luck is incomplete. In particular, Pritchard's definition of veritic luck takes into account only one type of epistemic risk (the risk of forming a false belief), but there are more senses of epistemic risk that define what is for a belief to be luckily true: the risk of believing closely related false propositions or the risk of believing a proposition with gappy content, paradoxical content or with no content at all.

35 Answer(s) to be found in section 7.6.2.

36 Although bracketing skeptical hypotheses when defining knowledge is common in epistemology, not all epistemologists make this move. For example, the main motivation of Robert Nozick's theory of knowledge (1981) is, precisely, the exclusion of skeptical hypotheses.

Crucial to my discussion was the point that these three forms of epistemic risk do not exhaust the sense in which a belief might be true by luck. I distinguished another type of risk that concerns several attributes of the agent and not merely the modal profile of her beliefs (I called it agent-focused epistemic risk) and I argued that it may be accommodated into an account of epistemic luck in the form of an epistemic analogue of the lack of control condition for luck. The relevant notion of epistemic control was left undefined, because the aim was simply to distinguish a further sense in which a belief might be true by luck. Accordingly, I made a distinction between two technical notions: veritic luck and veritic fortune. I then advanced an overarching hypothesis: these two epistemic phenomena exhaust the class of true beliefs that are not knowledge. I then qualified the hypothesis and claimed that there might be some cases of not known true belief in which none of these phenomena is instantiated. In any case, we can conclude by stating the core working hypothesis of the rest of the dissertation: *knowledge requires epistemic control*. The next chapter will provide the first steps towards an account of epistemic control.

5 MODAL EPISTEMOLOGY: EPISTEMIC CONTROL AS TRACKING

Chapter 4 introduced the hypothesis that nearly all (if not all) cases of true belief that is not knowledge are cases either of veritic luck or veritic fortune and the key idea was that both veritic luck and veritic fortune arise when an agent lacks epistemic control. The chapter ended with the corresponding working hypothesis that knowledge requires epistemic control. In chapter 3, I distinguished two complementary forms of control to which we refer when we use the term 'control' in ordinary contexts. *Effective control*, on the one hand, is roughly the kind of control that *A* has over *B* when *A* has the disposition and the aim to cause *B* to be in certain state and *B* is in that state because of *A*'s disposition. *Tracking control*, on the other hand, arises when *A* monitors *B*. With the purpose of translating and applying these two complementary varieties of control to the epistemic case, I will review two leading approaches to knowledge that are particularly suitable to this aim: modal and virtue epistemology. We will see, in the course of the discussion, that each approach captures one dimension of the kind of control that knowledge requires. In particular, some of the conditions that modal and virtue epistemologists put forward as necessary conditions for knowledge will allow to make sense of the notion of epistemic control.

Before entering into details, let me briefly present the general line of argumentation that motivates all the evaluations of modal and virtue-theoretic accounts of knowledge that will be made: I will defend that knowledge is a kind of cognitive achievement (as virtue epistemologists maintain) but (unlike them) I will define the notion of cognitive achievement in terms of the notion of epistemic control. In my view, knowledge arises just in case the agent has epistemic control over her cognitive performance. Crucially, since modal epistemology captures one substantial aspect of the notion of epistemic control, my proposal of understanding the notion of cognitive achievement in terms of epistemic control and knowledge in terms of cognitive achievement will allow to unify the modal and virtue-theoretic approaches to knowledge into one general and comprehensive theoretical framework. In chapter 6, I will present the argument in more detail.

The present chapter argues that modal epistemology is especially well suited to account for an epistemic analogue of the notion of tracking control and thus to make sense of the idea that knowledge requires some sort of monitoring.[1] I will start by analyzing the metaphor that motivated Nozick's account of knowledge: that knowledge is a matter of tracking the truth. I will then explain how Nozick's notion of tracking the truth can be seen as an instance of the kind of monitoring required for tracking control. The discussion will show that the typical epistemic conditions that modal epistemologists put forward (sensitivity, receptivity, safety) constitute different ways of conceptualizing the notion of epistemic control. Each condition will be analyzed separately and its pros and cons will be reviewed. The conclusion will be that a safety condition is the most adequate way to conceptualize the modal dimension of the notion of epistemic control and, in turn, the concept of knowledge. Nonetheless, a full defense of safety will have to wait until chapter 9, where I will defend it from some recent attacks against its necessity for knowledge.

5.1 NOZICK'S TRACKING METAPHOR

Modal epistemology is fundamentally guided by the idea that knowledge consists in *tracking the truth*. It was Robert Nozick who originally coined the metaphor:

> To know is to have a belief that tracks the truth. Knowledge is a particular way of being connected to the world, having a specific real factual connection to the world: tracking it. (Nozick 1981: 178)

When epistemologists address Nozick's theory of knowledge (1981) they tend to go to the specifics of the theory without paying much attention to the general picture of knowledge that motivates the account. Particular attention is usually paid to Nozick's epistemic conditions (the *sensitivity* and the *receptivity* conditions) and the consequences they carry for important epistemological issues such as epistemic closure or the skeptical problem. This is perhaps due to the fact that Nozick himself pays a lot of attention to the details of his definition of knowledge and is mainly concerned with giving a solution to the skeptical problem that exploits the thesis that knowledge is not closed under known entailment, a thesis that his analysis of knowledge implies. However, at some points, Nozick raises the level of generality of his discussion in order to motivate the general picture of knowing as tracking. It is this general picture that is of particular interest to the purpose of connecting the concept of knowledge with the notion of control. Let us see then how Nozick motivates his account.

§ *First motivation: the thought experiment.* Nozick is particularly concerned with the question of how people are to reach the state of hav-

1 Chapter 7 is devoted to the agential dimension of the notion of epistemic control, and chapters 8 and 9 examine how the agential and the modal dimensions are related.

ing true beliefs in a changing world. His answer takes the form of a though experiment:

> Imagine yourself in the position of God wanting to create organisms who would have (merely) true beliefs in a diverse and changing world. The ways to accomplish this are to (a) constantly intervene: start them with true beliefs and intervene to change their beliefs each time the world changes; (b) determine what the whole future will be, and create each being with a preprogrammed sequence of beliefs to fit his changing situation in a preestablished harmony; (c) create beings able to detect changes in facts who will change their beliefs accordingly. (Nozick 1981: 283-4)

According to our ordinary intuitions about knowledge, what type of situation is most plausible? Let us analyze (a). Typical cases of veritic fortune are cases in which the world is changed so that it fits the contents of one's beliefs. In a world like the one described in (a), something more radical happens: what is changed are the beliefs themselves so as to fit the world as the world changes. The problem then is not that the relevant beliefs are inherently fortunate (because the world changes to fit their contents). The problem is rather that there is no epistemic agency. If all beliefs are held and changed because of the intervention of an external being, how can inquiry be even possible? Inquiry is a necessary component of what it means to be an epistemic agent, but how can one carry out some inquiry if one cannot even change one's mind by one's own means (if any)? Thus, to claim that there is knowledge in a world as described in (a) would be to maintain that knowledge is possible without inquiry, in particular, and without epistemic agency, in general.

On the other hand, only a hard determinist (an incompatibilist who holds that causal determinism is true and who rejects free will as a a consequence) would be happy to claim that our world is as described in (b) and hence that our knowledge of the external world (and even of ourselves) is possible because our true beliefs are part of a preprogrammed sequence that corresponds to a preestablished course of events. That is, on this view each of us would have assigned from birth a set of truths and falsehoods that would have to be believed during the course of one's life. However, this kind of stance, which is very similar to the skeptic's, is to be rejected, as it calls into question important ordinary intuitions about the concept of knowledge, such as the intuition that we are responsible for many of our beliefs: since there is nothing we can do but to believe as a matter of natural necessity what we are given to believe, we are not responsible for our beliefs.

In view of this, the most plausible and simple explanation of how we manage to acquire true beliefs in a changing world is option (c): we are endowed with cognitive capabilities that allow us to detect (i.e., to track) changes in the world. As Nozick further explains, this is highly compatible with the evolutionary history of human life.

§ Second motivation: the evolutionary explanation. Nozick offers an evolutionary explanation of why we came to have tracking capabilities. The explanation begins with the plausible assumption that possessing the biological preconditions of belief is an adaptive advantage, connected with the advantages of having true beliefs (Nozick 1981: 284). The explanation continues as follows:

> The evolutionary process can give organisms true beliefs (in a changing world) only by giving them the capability to have true beliefs: so, it will give them more than (merely) true beliefs. In giving them a capability for true beliefs, it makes their beliefs (sometimes) vary somehow with the truth of what is believed; it makes their beliefs somehow sensitive to the facts. However, the evolutionary process will not bestow upon them a capability for true beliefs so powerful that in no logically possible situation would their beliefs be mistaken. Even if such capacity could arise by random mutation, (...) there would not be strong selection for it; there would be no selection against those other organisms whose lesser capacities function just as well in the actual range of situations. (Nozick 1981: 285)

If the thought experiment above motivates the idea that, in the light of our intuitions about knowledge, the best explanation of how we manage to acquire true beliefs in a changing world is the fact that we have tracking capacities, this evolutionary explanation shows that these tracking capacities give rise to knowledge by making human beliefs vary with the truth of what is believed. In addition, it shows that we do not need to monitor any possible variation, but only variations that are relevant to the actual range of situations. After all, as Nozick argues, organisms with the capacity to monitor any possible variation in any possible environment would be as well adapted to a certain environment as organisms with the more limited capacity to track the changes that take place just in *that* environment.

5.2 MONITORING, TRACKING AND FORCING

In chapter 3, we saw that many times we would not say that A controls B unless A has tracking control over B, i.e., unless A monitors B in some way. On other occasions, it suffices that A monitors B to claim that A controls B. On the other hand, we saw that monitoring might be merely informational (when A simply compiles information about B) and dispositional (when A, by compiling information about B, puts herself or itself in a position to act in a way that will eventually allow

A to achieve some of her or its aims, i.e., when *A*'s information about *B* disposes *A* to take certain courses of action that might lead *A* to the accomplishment of some of her/it goals). I gave several examples of dispositional monitoring: the astronaut who periodically consults the control panel of the spacecraft to check whether the speed and trajectory are as the NASA engineers planned (and to take proper action if not); the doctor who runs tests on her patient to decide whether to apply certain treatment; the thermostat that records temperature changes and activates or deactivates the heating as a consequence. In this section, I will explore how these ideas can be applied to the epistemic case.

Samuel F. Barker (1987), who is especially interested in the reach of Nozick's tracking metaphor, explains that its origin comes from the field of hunting: hunters literally track their quarries. The verb 'to track', however, is used in more metaphorical ways: missile systems are said to track targets and microwave antennas to track satellites. Barker thinks that all these examples show that tracking "whether it occurs in a literal or metaphorical sense, involves movement by the tracker in close response to the movement of its target" (Barker 1987: 284).

More importantly, all these cases exemplify forms of dispositional monitoring. Dispositional monitoring consists in gathering information to initiate, continue or stop some action or set of actions needed to accomplish some goal: the hunter carefully looks for the tracks of her quarry so as to move to a position where she can kill it; the radar of the missile system detects targets in the vicinity so that the missile launcher can be properly aimed; the microwave antenna detects the passing satellite so as to rotate to a position where the signal strength is maximal. As we see, all the trackers in question have a goal (killing the quarrel, hitting the targets, maximizing the signal strength) and all of them take proper action (walking to certain direction, aiming the missile launcher to the targets, rotating) in virtue of some instance of monitoring or tracking. The epistemic case is not much different. Barker asks:

> What is the notion of tracking supposed to mean, however, if we stretch it considerably further to cover a relationship between belief and the truth, where change of place is not in question? (Barker 1987: 284)

And he answers:

> Here Nozick's suggestive idea is that we may speak of tracking when a person's state of belief about some matter reliably adjusts itself to conform to the truth about that matter. This person's state of belief does and would change to suit varying situations, so that correspondence is and would be maintained between what is believed and

> what is the case. In this sense, the person's state of belief is analogous to the antenna which rotates so as to stay locked in on its target: each is 'tracking' something, in that each alters as is required in order to maintain a certain conformity between itself and what it, so to speak, tracks. (*ibid.*)

My contention is that the *modal dimension* of the kind of control that knowledge requires consists in our state of belief changing to suit varying hypothetical situations. As we will see in section 7.6.1, it is legitimate to talk about non-voluntary control over belief insofar as we are able to change our minds and thus form or revise our doxastic attitudes. When we change our minds in this way, we put ourselves in a better position to form or revise other beliefs. More specifically, the kind of mental actions that this sort of doxastic monitoring disposes us to take would be to initiate or to stop belief formation and to continue or to stop belief maintenance so that conformity is maintained between what is believed and what is and what would be the case. Less specifically, a belief that tracks some truth puts us in a position to track some other truths. Nozick's tracking the truth metaphor can be thus understood as being about a special type of dispositional monitoring.

When it comes to defining knowledge, the obvious way to specify how a person's state of belief should change in order for her belief to qualify as knowledge is the way Nozick actually followed. Nozick thought that the idea of tracking the truth was to be specified through subjunctive conditionals that restrict the range of possible scenarios in which the agent would believe truths or falsehoods. More generally, this is what Vincent F. Hendricks (2006) calls the *forcing strategy*:

> The idea of forcing is to delimit the set of possibilities over which the inquiring agent has to succeed: If the agent can succeed over the relevant possibility set, then the agent may still be said to have knowledge even if he commits many errors, even grave ones, in other but irrelevant possibilities. (Hendricks 2006: 10)

The modal dimension of epistemic control is, therefore, a restricted one: an agent is said to have control over her cognitive performance, i.e., of her coming to believe some proposition p, only if she would believe the truth about p (or would not believe that p falsely) in a certain range of hypothetical scenarios. Difficulties arise, however, when it comes to delimiting the concrete set of possibilities over which the agent has to succeed or not fail. Under which conditions does an agent S track the truth as regards a proposition p? Nozick experimented with different forcing conditions that he had to drop because they succumbed to several problems,[2] after which he concluded that when *S* knows that *p* the following two conditions must be true of *S*:

- *Sensitivity*: If *p* were not true, *S* would not believe that *p*.[3]
- *Receptivity*: If *p* were true, *S* would believe that *p*.[4]

Sensitivity and receptivity are not the only conditions that account for the idea of tracking, i.e., they are not the only way of delimiting the set of possibilities over which the inquiring agent has to succeed or not fail. Sosa (1999), dissatisfied with the consequences of Nozick's analysis of knowledge, introduced the following modal condition:

- *Safety*: If *S* were to believe that *p*, *p* would be true.

I will review the problems and advantages of each condition in the next three sections. For the moment, the discussion has shown that the notion of tracking control, i.e., of monitoring, corresponds in the epistemic case to Nozick's idea of tracking the truth, an idea that may be implemented using modal conditions for knowledge that specify how conformity should be maintained between what is believed and what is and what would be the case. It is in this way in which the modal dimension of epistemic control should be understood.

Finally, note that the notions of epistemic luck and epistemic fortune have a modal dimension as well. More specifically, we saw in chapter 4 that *S*'s true belief that *p* is veritically lucky only if in at least half the close possible worlds in which *S* believes that *p*, *p* is false; and veritically fortunate only if in more than half the close possible worlds in which *S* believes that *p*, *p* is true. Accordingly, in cases of *veritic luck* the agent's actual state of belief would vary in such a way that correspondence would not be maintained between what is believed (*p*) and what is the case in a certain range of close possible worlds (~ *p*). By contrast, in cases of veritic fortune such a correspondence is preserved.[5] Accordingly, it seems then that an appropriate forcing condition for knowledge, i.e., a condition that states in which modal sense knowledge requires epistemic control, should be able to rule out at least cases of veritic luck as cases of knowledge.

2 See Nozick (1981: 682, fn. 12). For an excellent discussion on epistemic modal conditions, see Égré (2008).

3 Nozick calls this condition *variance* and calls sensitivity the conjunction of variance and receptivity. In keeping with most commentators, I use the term 'sensitivity' to refer only to the variance condition.

4 Nozick calls this condition *adherence*. The term 'receptivity' is from Sosa (2002).

5 There might be borderline cases: cases in which it is not clear whether a true belief instantiates veritic luck or veritic fortune, either because the case is under-described so that we do not know the proportion of close possible worlds in which the target belief would fail, or because the target belief would be false in approximately half the close possible worlds.

5.3 SENSITIVITY

5.3.1 *Pros*

The sensitivity condition has a relative success when it comes to the exclusion of veritic luck and veritic fortune. In general, it yields good results when explaining Gettier-style cases. For example, it rules out cases of *veritic intervening luck* such as case F (the snake in the tree example), which is a paradigmatic Gettier-style case, as cases of knowledge. Adam's true belief that there is a snake in the tree is not sensitive, because if it were not true that there is a snake in the tree (in the hypothetical case that the snake does not decide to climb the tree), then Adam would still believe that there is snake in the tree (because in that case he would still keep looking at the snake-looking branch).[6]

On the other hand, sensitivity can rule out cases of *veritic environmental luck* such as FAKE BARNS, which is another Gettier-style case (or so is typically considered), as cases of knowledge. Henry's belief that the object in front of him is a barn is not sensitive, because if it were not true that the object in front of him is a barn (in the hypothetical case that he looks at a fake barn), then he would still believe that it is a barn (because the fake is an indistinguishable fake). Therefore, Henry's belief is not sensitive.

In addition, sensitivity is able to rule out *some* cases of *veritic fortune* as cases of knowledge. Recall TEMP and STILL SNAKE. Suppose that Temp, by looking at the broken thermometer, forms the belief that the temperature is 25ºC and that this belief is true because, unbeknownst to Temp, the hidden agent has raised the temperature of the room from 24ºC to 25ºC. If it were not true that the temperature is 25ºC (in the hypothetical case that the hidden agent does not intervene), Temp would continue to believe it (because he would form his belief by looking at the broken thermometer, which would read 25ºC). In that case, Temp's belief would be insensitive. On the other hand, the sensitivity condition delivers the correct judgment of STILL SNAKE as well: as in case F, if there were no snake in the tree (something that would not occur in any close possible world because it is a stipulation of the case that the snake has always lived in that tree and has never come down), Adam would still believe that there is a snake in the tree (because he would still keep looking at the snake-looking branch).

5.3.2 *Cons*

There is a sort of consensus among epistemologists that Nozick's analysis of knowledge, although a laudable effort that contributed to a new way of thinking about knowledge, is not a correct analy-

6 The same reasoning applies to STOPPED CLOCK.

sis of the concept.[7] Here I am not interested in Nozick's definition of knowledge but in its two modal conditions separately (sensitivity and receptivity), because each condition specifies one way in which the modal dimension of epistemic control could be understood. The idea is to consider all the pros and cons of each condition separately and then choose the one that best serve our purposes.

§ *Sensitivity is compatible with veritic fortune.* First of all, let me note that sensitivity is not an adequate anti-fortune condition, as it cannot rule out some cases of veritic fortune as cases of knowledge.[8] In MARY, for example, if Mary's husband had not been in the room, she would not have believed that he is because the evil husband's counterpart would not have appeared before her eyes. Therefore, Mary's true belief that her husband is in the room is sensitive. Moreover, we can modify the other cases of veritic fortune considered in such a way that the target beliefs of the cases are sensitive but not known. By way of illustration, consider the following modification of TEMP:

> EXPLOSIVE TEMP
>
> As in the original case, the broken thermometer reads 25ºC and Temp's belief that the temperature is 25ºC is true because of the intervention of the hidden agent. However, unlike in the original case, if the hidden agent does not adjust the temperature of the room to match the reading of the thermometer, a bomb explodes drastically raising the temperature of the room.

Possible worlds in which it is not true that the temperature in the room is 25ºC are worlds in which the hidden agent does not intervene, the bomb explodes, the temperature drastically rises, Temp dies and hence does not falsely believe that the temperature is 25ºC. Therefore, in EXPLOSIVE TEMP Temp's belief is sensitive: if it were false that the temperature is 25ºC, Temp would not believe it.[9]

§ *Sensitivity violates epistemic closure*. Many philosophers think that no adequate theory of knowledge should entail the negation of the principle of closure of knowledge under known entailment (roughly, the principle that if S knows that p, and knows that p entails q, then S knows that q). However, sensitivity violates epistemic closure. To see this, suppose that I look at my hands (in the actual world) and I come

7 In a remarkable but, in my opinion, quite fruitless attempt, some commentators have tried to resurrect sensitivity-based theories (see especially Adams & Clarke 2005). More promising is Sherrilyn Roush's project (2005) of reformulating sensitivity in probabilistic terms.

8 As we will see, neither receptivity nor safety are anti-fortune conditions. Consequently, Nozick's account of knowledge as sensitive and receptive true belief is not sufficient for knowledge.

9 Similar modifications can be made to STILL SNAKE.

to believe that I have hands. My belief that I have hands is sensitive: if I did not have hands (because, say, I lost them in an accident), I would not believe that I have hands. Presumably (when the rest of epistemic conditions are in place), that I have hands is something that I know. On the other hand, my true belief that I am not a brain in a vat is not sensitive: if I were a brain in a vat, I would still believe that I am not (because evil neuroscientists would make me believe it). Therefore, if sensitivity is necessary for knowledge, I do not know that I am not a brain in a vat (sensitivity does not hold). Yet, that I have hands entails that I am not a brain in a vat. Moreover, I know this entailment, but I do not know its conclusion (my belief in the conclusion is not sensitive). Therefore, sensitivity entails that knowledge is not closed under known entailment.

The rejection of epistemic closure carries undesired consequences for sensitivity-based theories: it leads to what Keith DeRose (1995) has called 'abominable conjunctions'. As we have just seen, Nozick's analysis of knowledge (knowledge is sensitive and receptive true belief) entails that I know that I have hands *and* that I do not know that I am a brain in a vat. This is acceptable to some extent and Nozick's response to the skeptic lies precisely in this conjunction. However, some philosophers have found this point highly problematic. At any rate, the main reason not to specify the modal dimension of epistemic control in terms of sensitivity comes next.

§ Sensitivity is not necessary for knowledge. Counterexamples have been constructed to show that sensitivity is not necessary for knowledge. As Jonathan Vogel and Ernest Sosa have insisted in several places, the condition offers the wrong results in cases of inductive knowledge. Consider the following two examples:

> X-Ray
>
> Roger places a piece of uranium on a photographic plate, and discovers that the plate has become fogged. He repeats the experiment many times. After numerous trials, he puts a piece of uranium on a plate, goes away from his laboratory, and returns some time later. Roger believes that the plate is fogged. Moreover he knows, by induction, that the plate is fogged, even before he inspects it. (Vogel 2007: 78)
>
> Trash Bag
>
> On John's way to the elevator he releases a trash bag down the chute from his high rise condo. Presumably, he knows his bag will soon be in the basement. (Adapted from Sosa 1999)

Unless we are highly skeptic about induction we have to admit that there is knowledge in these cases, despite the fact, however, that the target beliefs are not sensitive. In X-Ray, if the plate were not fogged (because of some odd occurrence), Roger would believe, before inspecting it, that the plate is fogged. In Trash Bag, if the bag were snagged somehow in the chute (something that would not occur in any close possible world), John would still believe that it is in the basement. Sensitivity, therefore, is not a necessary condition for knowledge, which is a good reason why the condition does not serve to specify the modal dimension of the notion of epistemic control.

5.4 RECEPTIVITY

5.4.1 *Pros*

Nozick is aware that a true belief can be sensitive without being knowledge. For example, if you are a brain in a vat being electrochemically stimulated to believe that you are a brain in a vat, sensitivity holds: you would not believe that you are a brain in a vat if you were not (if you were not, you would have properly functioning senses, and so you would believe that you are not a brain in a vat, as we ordinarily do).

According to Nozick, receptivity rules out this kind of brain-in-a-vat scenarios as cases of knowledge because when one is a brain in a vat there are close possible worlds in which one is electrochemically stimulated to believe that one is not a brain in a vat.[10] That is, contrary to what receptivity says, had you been a brain in a vat, you would have not believed it.[11] In addition, Nozick thinks that the condition allows to rule out the following much-discussed case as a case of knowledge:

> Political Leader
>
> A political leader is assassinated. (...) On nationwide television it is announced that an assassination attempt has failed to kill the leader. Before the announcement is made,

10 Nozick simply stipulates here that in brain-in-a-vat scenarios one could easily be electrochemically stimulated to believe that one is *not* a brain in a vat. But, of course, this is only one among the several possible stipulations that one could make about this kind of scenarios. Other stipulations could entail that, while envated, one could easily be electrochemically stimulated to believe that one is a brain in a vat. Therefore, when it comes to skeptical cases the success of receptivity is contentious.

11 Nozick sketches a modification of Lewis/Stalnaker's semantics of subjunctive conditionals to validate the assumption that receptivity is not trivially true whenever the antecedent and the consequent are true in the actual world, i.e., when S truly believes that p in the actual world. On the Lewis/Stalnaker's semantics, when A is true, the subjunctive conditional if A were true, B would be true ($A > B$) is true if and only if B is also true (for further discussion see section 1.7 of Lewis's monograph [1973]). As per Nozick's modification, if A is true then $A > B$ is true if and only if B is true in all A-worlds closer (by the metric) to the actual world than is any not-A-world (Nozick sketches this proposal in a footnote of his book [1981: 680-1, fn. 8])

> a newspaper publishes the real story. For some improbable reason, Jill does not watch the television, but reads the newspaper. What she reads is true but she does not know that the political leader has been assassinated. (Adapted from Harman 1973: 142-3)

It is not so clear that there is no knowledge in this case. However, Nozick shares Harman's intuition that Jill does not know.[12] Jill's belief is sensitive because, if the political leader had not been assassinated, she would not have believed it, because the newspaper would not have reported the assassination. However, Jill's belief is not receptive: in possible scenarios where the political leader is assassinated (scenarios different from the actual one) Jill watches the television announcement that the assassination attempt has failed and believes falsely that the leader is alive. Therefore, receptivity does not hold: had the leader been assassinated, Jill would not have believed it.

5.4.2 *Cons*

§ Receptivity is compatible with veritic fortune. Consider MARY again. In possible worlds in which the proposition <my husband is in the room> is true, Mary continues to believe that her husband is in the room. Since MARY is a case of veritic fortune and receptivity is satisfied in it, receptivity is compatible with veritic fortune.

§ Tracking à la Nozick: more abominable conjunctions. Let me make a brief excursus on Nozick's analysis of knowledge (not just on receptivity). Nozick's *coup de grâce* comes from Saul Kripke, who shows that sensitivity and receptivity in conjunction lead to highly controversial consequences.[13] Consider again POLITICAL LEADER. Kripke (2011: 181-2) argues that Jill not only believes that (*i*) the political leader has died but also that (*ii*) she has read an uncontradicted newspaper report to this effect.

Jill's belief in (*i*) and (*ii*) is sensitive: had the conjunction not been true, she would not have believed it, as in that case no newspaper would have published a report informing of the political leader's death ($\sim (ii)$), because the leader would have been alive ($\sim (i)$).

Kripke's point is that Jill's belief in (*i*) and (*ii*) is not only sensitive but also receptive. To justify this point, Kripke argues that, when assessing the belief that (*i*) and (*ii*) for receptivity, (*ii*) guarantees that we focus on possible worlds in which Jill hears no denial (e.g., possible worlds in which the nationwide television falsely announces that

12 In my opinion, this intuition is incorrect: one might come to know p via method M_1 (e.g., reading a newspaper) even though one could easily have come to believe $\sim p$ via M_2 (e.g., watching the news on TV).

13 The reason to consider Kripke's objections to Nozick's analysis is to show that the conjunction of sensitivity and receptivity is not a promising way of specifying the modal dimension of epistemic control.

the leader has survived but in which Jill does not switch on the television), Jill continues to believe that the leader is dead.

If Kripke's is right on this point, then Nozick's analysis (knowledge as sensitive and receptive true belief) predicts that Jill knows that the political leader has not died *and* that she has heard an uncontradicted newspaper report to this effect. What is the problem? The problem is that, as we have seen above, Jill's belief that the leader is death is *not* receptive: again, in possible scenarios where the political leader is assassinated (scenarios different from the actual one) Jill watches the television announcement that the assassination attempt has failed and believes falsely that the leader is alive. Given Nozick's definition of knowledge, Jill does not know that (*i*) the political is death despite the fact that she knows that (*i*) the political leader is death *and* that (*ii*) she has read an uncontradicted newspaper report to this effect.

As far as I can see, the only possible reply that the Nozickian could try would be to dispute that when assessing Jill's belief for receptivity our attention is solely confined to counterfactual situations in which she hears no denial, as Kripke maintains.

However, such an effort would be useless to save Nozick's analysis, since Kripke (2011: 186) shows that sensitivity leads to equally uncomfortable conjunctions.[14] Consider again Red Barn. Carl looks at the only real barn in the area (the barn is red and the area is populated with yellow fake barns). Suppose that he comes to believe that the object in front of him is a barn. Carl's belief is not sensitive, because if there had been no barn in front of him, he would still have believed it (because he would have looked at some fake barn). However, this time Carl forms the true belief that the object in front of him is a *red* barn. Kripke's point is that Carl's belief is sensitive because had not been a red barn in front of him, there would have been a *yellow* fake barn, and hence he would *not* have formed the belief that the object in front of him is a red barn. Since receptivity also holds (if there were a red barn, Carl would have believed that it was), Nozick's analysis entails the unwelcome result that one might not know that the object in front of one is a barn while knowing that it is a red barn.

In conclusion, the conjunction of receptivity and sensitivity is not a very promising way of specifying the modal dimension of the notion of epistemic control. Could that dimension be specified solely in terms of receptivity? There is a powerful reason not to.

§ *Receptivity is not necessary for knowledge.* Sosa shows that the requirement is too strong. Consider the following case:

> Birds
>
> There is a large pelican on the lawn in plain daylight that could fly away at any time. Michelle looks at it and forms

14 Kripke's paper contains so many objections to Nozick's account of knowledge that it has come to be known as the *Nozick-bashing lectures*.

> the belief that there is a bird on the lawn. There is also a small robin perched in the shade, unseen to Michelle. (Adapted from Sosa 2002)

Michelle's belief is sensitive because had no bird been on the lawn (neither the pelican nor the robin), she would not have believed that there is a bird on the lawn. However, her belief is not receptive. Let the antecedent of receptivity be true: imagine possible worlds in which there is a bird on the lawn. The possibilities are the following: 1) both the pelican and the robin are on the lawn; 2) only the pelican is on the lawn; 3) only the robin is on the lawn. When the antecedent of receptivity is true in virtue of the first or the second possibilities, the consequent is also true: Michelle believes that there is a bird on the lawn. However, when the antecedent is true in virtue of the third possibility, the consequent is not true: Michelle does not believe that there is a bird on the lawn because the robin is too small to see it. Yet, it seems that Michelle *knows* that there is a bird on the lawn when she looks at the pelican in the actual world. Conclusion: receptivity is not necessary for knowledge.

We are trying to make sense of (the modal dimension of) epistemic control, which by hypothesis is *required* for knowledge. Since neither sensitivity nor receptivity are necessary for knowledge we cannot expect to specify this modal dimension in terms of these modal conditions. We need, therefore, to explore further alternatives. The most promising candidate is the safety principle.

5.5 SAFETY

One knows a proposition only if one's belief could not *easily* have been false. This is, roughly, the core idea underlying the safety principle. As we have seen, Sosa's version (1999) takes the form of a subjunctive conditional:

- SAFETY$_S$: If S knows that p, then if S were to believe that p, p would be true. $[Kp \rightarrow (Bp > p)]$

On the other hand, Williamson (2000) appeals to standards of similarity to define safety. It should be noted that Williamson does not define knowledge in terms of a safety condition because it is his position that knowledge is not analyzable. But he does consider it a necessary condition for knowledge. How is that? Williamson is particularly interested in the notion of reliability, which he does regard as necessary for knowledge and which he takes to be broad, to the extent that "[r]eliability resembles safety, stability, and robustness" (2000: 124). He claims (2000: 100): "No reason has emerged to doubt the intuitive claim that reliability is necessary for knowledge". Contrary to what one might think, it does not follow from this claim that knowledge is analyzable, since, as Williamson points out, it may be impossible to frame other necessary conditions whose conjunction with reliability is a necessary and sufficient condition for knowledge. This is how Williamson takes safety to be necessary for knowledge without being committed to analyze knowledge in terms of safety.

According to Paul Égré (2008), Williamson's reliability entails two dual technical notions: safety and robustness. On Égré's characterization, Williamson's safety is mere avoidance for false belief in similar cases. This is supported by some textual evidence: "If one believes p truly in a case α, one must avoid false belief in other cases sufficiently similar to α in order to count as reliable enough to know p in α." (Williamson 2000: 100). In a formal way:

- SAFETY$_w$: $Kp \rightarrow \Box(p \rightarrow \sim B \sim p)$[15][16]

Finally, Pritchard (2005) formulates the safety principle explicitly in terms of possible worlds. Here is his version:

- SAFETY I: If S knows that p, then in nearly all (if not all) close possible worlds in which she forms her belief about p in the same way as in the actual world, S only believes that p when is true p.

In general, we see that the kind of 'forcing' that the safety condition carries out consists in restricting the range of counterfactual situations, possible worlds or cases in which the agent has to succeed (or not fail) to those that are close, nearby or similar to the actual situation, world or case. Why should (some version of) safety be part of our theory of knowledge?

5.5.1 *Pros*

§ *Safety and veritic luck.* The first reason is that safety is extremely effective against veritic luck or, in other words, against any form of knowledge-undermining belief-focused risk. To see this, let us consider some paradigmatic cases of veritic luck (I will use SAFETY I for discussion). Let us begin with a case of veritic *intervening luck.* Consider case F (the Gettier-style case of the snake in the tree). Adam does not know that there is a snake in the tree because his belief is risky in the first sense of belief-focused risk (BELIEF-FOCUSED RISK I). The consequent of SAFETY I does not hold, because in most close possible

15 The necessity operator rules over a range of similar cases. This is compatible with Williamson's more recent remarks that safety is a sort of "local necessity" (Williamson 2009), an idea that was discussed in chapter 1.

16 Égré compares this negative notion of safety as avoidance of error with the notion of robustness as propensity of true belief for similar cases, which he defines as follows: $Kp \rightarrow \Box(p \rightarrow Bp)$. As Égré acknowledges, this condition, which amounts to Nozick's receptivity, is too strong for knowledge (there are cases of unrobust known belief).

worlds in which he continues to believe that there is snake in the tree and the relevant initial conditions for that belief are the same as in the actual world (i.e., close possible worlds in which he looks at the snake-looking branch), the snake does not climb the tree and hence his belief is false.

Does safety rule out cases of *environmental luck* as cases of knowledge? Consider FAKE BARNS. Henry looks at the only real barn in the area and comes to believe that there is a barn in front of him. In close possible worlds in which he continues to believe the same proposition and in which the relevant initial conditions for that belief are the same as in the actual world (worlds in which he looks at objects with the visible characteristics of real barns), Henry looks at barn replicas and thus comes to believe *falsely* that there is a barn in front of him. Thus, as per SAFETY I, Henry's belief is not knowledge. His belief is unsafe.

What about the other senses of knowledge-undermining belief-focused risk, e.g., cases in which there is no risk of error but of *close error*? Consider RED BARN. Carl looks at the only real barn in the area and comes to believe that there is a red barn in front of him, which is true. In close possible worlds in which he looks at other objects with the visible features of a barn, he does not come to believe that there is a *red* barn in front of him falsely, because all barn replicas are yellow. The consequent of SAFETY I thus holds. Should we accordingly conclude that Carl's belief is safe?

As we will see next, the answer is negative. Sosa, Williamson and Pritchard's definitions of safety only state one sense in which a belief might be safe (and therefore only one sense in which it might be unsafe). What cases of knowledge-undermining risk as defined by BELIEF-FOCUSED RISK II show is that we need to extend the safety condition to cover any form of knowledge-undermining risk. I propose the following safety principle:

- SAFETY II: If S knows that p, then in nearly all (if not all) close possible worlds in which S forms beliefs in propositions $(p_1...p_n)$ closely related to p and in which the relevant initial conditions for those beliefs are the same as the initial conditions for her belief that p in the actual world, S only believes $(p_1...p_n)$ when $(p_1...p_n)$ are true.

Consider now RED BARN. In close possible worlds, Carl would believe that there is a yellow barn in front of him, which is false. Since the proposition that there is a yellow barn in front of him is closely related to the proposition that there is a red barn in front of him, the consequent of SAFETY II does not hold. Carl's belief is therefore unsafe.

What about the third sense of knowledge-undermining risk, i.e., the risk of believing a gappy proposition, a proposition with paradoxical content or with no content at all? As Manley (2007: 406) remarks, "it is tempting to reformulate the safety condition by simply requiring that the very thought token whereby *S* believes that *p* could not easily have been a failed thought", i.e., a thought with no content, false, gappy or paradoxical content. But as he correctly notes, this hangs too much on the individuation of thoughts (this caveat was taken into consideration in chapter 4 in the discussion on belief-focused risk). His proposal is to resort to a counterpart relation on thoughts, broader than identity, such that:

- SAFETY III: If *S* knows that *p*, then in nearly all (if not all) close possible worlds *S* does not form a counterpart belief with gappy content, paradoxical content or with no content all (were the relevant initial conditions for the counterpart belief be the same as the initial conditions for *S*'s belief that *p* in the actual world).[17]

An agent's true belief is *safe* just in case it meets the requirements of SAFETY I, II and III. The *safety condition* says then that *S* knows that *p*, only if *S*'s belief that *p* is safe in these three senses. Safety, so understood, is the anti-luck condition par excellence: no case of veritic luck, i.e., no case of knowledge-undermining belief-focused risk, is a case of safe belief. We have thus very good reason to use safety to account for the modal dimension of epistemic control and thus to include it to our theory of knowledge.

5.5.2 *Cons*

§ Safety is compatible with veritic fortune. While sensitivity can cope with some cases of veritic fortune (although not with all of them), safety completely fails in this regard. This should not be surprising. Safety is aimed at the exclusion of knowledge-undermining *belief-focused risk* but in cases of veritic fortune there is no significant close possibility of error. Automatically, this makes veritic fortunate beliefs safe.[18] In TEMP, for example, Temp's beliefs about the temperature are *guaranteed to be true* by the intervention of the hidden agent, who changes the room's temperature in order to match the readings of the

17 Manley's revised version of safety does not include a clause on relevant initial conditions: If *S* knows that *p*, *S* could not easily have had a failed counterpart thought (I have included one in SAFETY III). As I argued in chapter 4, there are cases of *knowledge* in which an agent could easily have had a failed counterpart thought. In those cases, close possible worlds in which the agent forms failed counterpart thoughts are worlds in which the initial conditions for those thoughts are not the same as in the actual world (this is roughly the reason why we ascribe knowledge to the agent).

18 This is not completely right, though. As we have defined the notion, there might be cases of veritic fortune in which the target beliefs are true in more than half the close possible worlds but not in nearly all, if not all, close possible worlds, so that the beliefs in question would be veritically fortunate but not safe. Nevertheless, in the cases of veritic fortune that we have been discussing so far there is no belief-focused risk involved and hence the target beliefs count as safe, as we will see next. For simplicity, I will confine myself to *paradigmatic* cases of veritic fortune.

broken thermometer. In MARY, the counterpart of Mary's husband appears before her eyes every time Mary is about to form the belief that her husband is in the room and only if her husband is, unseen to Mary, in the room, which *guarantees the truth* of Mary's beliefs in the whole range of close possible worlds. In STILL SNAKE, Adam forms the belief that there is snake in the tree and in no close possible world the snake is not there (it is an stipulation of the case that the snake has always lived in the tree and has never come down). Therefore, there is *no risk that Adam fails to believe the truth*. In sum, safety is not an anti-fortune condition.

§ Is safety necessary for knowledge? Cases of veritic fortune show that safety is *not sufficient* for knowledge. Juan Comesaña (2005), Christoph Kelp (2009) and Ram Neta and Guy Rohrbaugh (2004) have proposed several counterexamples with the aim of showing that safety is *not necessary* for knowledge either:

> HALLOWEEN PARTY
>
> There is a Halloween party at Andy's house, and I am invited. Andy's house is very difficult to find, so he hires Judy to stand at a crossroads and direct people towards the house (Judy's job is to tell people that the party is at the house down the left road). Unbeknownst to me, Andy doesn't want Michael to go to the party, so he also tells Judy that if she sees Michael she should tell him the same thing she tells everybody else (that the party is at the house down the left road), but she should immediately phone Andy so that the party can be moved to Adam's house, which is down the right road. I seriously consider disguising myself as Michael, but at the last moment I don't. When I get to the crossroads, I ask Judy where the party is, and she tells me that it is down the left road. (Comesaña 2005: 397)
>
> TOXIN
>
> I am drinking a glass of water which I have just poured from the bottle. Standing next to me is a happy person who has just won the lottery. Had this person lost the lottery, she would have maliciously polluted my water with a tasteless, odorless, colorless toxin. But since she won the lottery, she does no such thing. Nonetheless, she almost lost the lottery. Now, I drink the pure, unadulterated water and judge, truly and knowingly, that I am drinking pure, unadulterated water. But the toxin would not have flavored the water, and so had the toxin gone in, I would still have believed falsely that I was drinking pure, unadulterated water. (Neta & Rohrbaugh 2004: 399–400)

RUSSELL'S ARCH-NEMESIS

> Suppose Russell's arch-nemesis has an interest that Russell forms a belief (no matter whether true or not) that it's 8:22 by looking at the grandfather clock when he comes down the stairs. Russell's arch-nemesis is prepared to do whatever it may take in order to ensure that Russell acquires a belief that it's 8:22 by looking at the grandfather clock when he comes down the stairs. (...) However, Russell's arch-nemesis is also lazy. He will act only if Russell does not come down the stairs at 8:22 of his own accord. Suppose, as it so happens, Russell does come down the stairs at 8:22. Russell's arch-nemesis remains inactive. Russell forms a belief that it's 8:22. It is 8:22. (Kelp 2009: 27)

According to these authors, in all these examples 1) the agents have knowledge and 2) their beliefs are unsafe. In HALLOWEEN PARTY, Comesaña knows that the party is down the left road, but (allegedly) in close possible worlds where he believes that proposition in the same way, his belief is false. In RUSSELL'S ARCH-NEMESIS, Russell knows that it is 8:22, but (allegedly) that belief would have been false had Russell come to believe it in the same way. And the same (allegedly) applies to the belief that Neta is drinking pure unadulterated water in TOXIN. Do these examples prove that unsafe knowledge is possible?

The answer is negative. As Tomas Bogardus (2012) has recently argued, these cases are not counterexamples to safety because they locate the relevant epistemic risk *before* the agents form their beliefs, but it does not follow from that that their beliefs are unsafe. In HALLOWEEN PARTY, Comesaña was at risk of forming a false belief when deciding whether to disguise himself as Michael or not, but *once* he has decided not to he is no longer at risk of forming a false belief. Analogously, as Bogardus explains, in TOXIN I Neta was at risk of forming the false belief that he is drinking pure unadulterated water before the person next to him won the lottery, but *once* that person wins he is no longer at such a risk. Finally, in RUSSELL'S ARCH-NEMESIS Russell was at risk of forming the false belief that it is 8:22 when deciding when to come down the stairs, but *once* he decides to come down at 8:22 he is no longer at risk of forming a false belief.

Bogardus's point can be further supported by noting that the kind of mistake these authors make is the same mistake that Jennifer Lackey made in her counterexample to the necessity of the lack of control condition for luck that we discussed in chapter 3 (BURIED TREASURE). One of the lines of reply to Lackey's case (adopted by Coffman [2009]) was to show that the case commits her to the so-called *luck infection thesis*, the thesis that if it is by luck that S is positioned to φ, then it is by luck that S has φ-ed. But the thesis has blatant counterexamples. For instance, it can be by luck that Kobe Bryant has found a basketball to make a slam dunk, but once he has the basketball it is not by luck that he makes it. Analogously, from the fact that it is by luck that *S* is *in a position to* know that *p* it does not necessarily follow that *S* luckily knows that *p*. To put it differently: from the fact that *S* was at risk of not being in a position to know that *p* it does not necessarily follow that *S*'s belief that *p* is formed unsafely. Another way to phrase the point is this: luck structuring the circumstances (e.g., luck involved in how one comes to be in a position to know) is not inherited by the actions and events that take place in those circumstances (e.g., one's coming to believe some truth).

The way the safety principle accommodates these intuitions is by means of the initial conditions clause. In the alleged counterexamples to safety, possible worlds in which the agents in question form false beliefs are worlds in which the relevant initial conditions for their beliefs are not the same as in the actual world. In HALLOWEEN PARTY, Comesaña comes to believe that the party is down the left road by hearing Judy's sincere testimony. In possible worlds in which he *decides* to disguise himself as Michael, Judy is no longer sincere and Comesaña forms a false belief. Comesaña's decision of not disguising himself as Michael is part of the initial conditions for his actual belief. In TOXIN, Neta comes to believe that he is drinking pure, unadulterated pure water by drinking pure water that the lottery winner *decided not* to alter. The fact that that person decided so, is part of the relevant initial conditions for his belief (together with the exercise of his visual abilities).[19] The same reasoning applies to RUSSELL'S ARCH-NEMESIS: the relevant initial conditions for actual Russell's belief include the decision of his arch-nemesis of not stopping the clock. With all these considerations in mind, we can conclude that the cases proposed by these authors do not succeed in showing that safety is not necessary for knowledge.[20]

A more serious objection to safety comes from Bogardus himself. He thinks that a genuine counterexample to the necessity of the condition would be a case in which the agent in question is at epistemic risk *when* she forms the relevant belief and not merely *before*. He constructs, accordingly, an example of knowledge in which epistemic risk is located during belief formation. However, a full defense of the safety condition for knowledge (and hence an answer to Bogardus's counterexample) will have to wait until chapter 9. For the moment, and *pace* Bogardus, we can conclude that safety is necessary for knowledge.

19 A similar point was made in chapter 3 about whimsical events (events that result from decisions made on a whim).

20 In chapter 9, I will offer a principle for individuating methods of belief formation that will allow us to understand better why cases of this sort are not counterexamples to safety.

5.6 Safety and (epistemic) control

When a driver controls her car, not easily could she have an accident while driving in the way she does. In a sense, it is not by luck that she does not have an accident. Her driving and her success (not having an accident) are said to be safe. Generalizing, it seems then that one of the reasons (if not the main reason) why control is incompatible with luck is that safety is a necessary condition for control.

Turning now to the epistemic case, we have seen that the notion of epistemic control can be nicely conceptualized in terms of an agent's belief on some matter reliably adjusting itself to conform to the truth about that matter in a range of hypothetical situations. This is the epistemic counterpart of the notion of tracking control. In which range of hypothetical situations should an agent succeed or not fail? I have analyzed three proposals, three different forcing conditions: sensitivity, receptivity and safety. There is no other option but to rule out receptivity and sensitivity as necessary for epistemic control, precisely because they are not necessary for knowledge and, by hypothesis, knowledge requires epistemic control.

Pace Bogardus, safety is necessary for knowledge. In addition, it delimits well the possibilities over which an agent should succeed in order to gain knowledge, not only because it excludes veritic luck, but also because it excludes it without carrying undesired consequences such as the violation of epistemic closure or the implication of abominable conjunctions. Accordingly, and also in keeping with the previous considerations on the relation between the general notions of safety and control, I anticipate the first condition in the analysis of the notion of epistemic control:

- Safety: *S* controls her cognitive performance only if her cognitive success is safe.

In the next chapter, I will explain what is meant by cognitive performance and cognitive success. Before that, let us recap.

5.7 Summary

This chapter started with the hypothesis that knowledge requires epistemic control and its aim has been to find an epistemic analogue of the notion of tracking control. In sections 5.1 and 5.2., I have introduced Robert Nozick's fruitful idea that knowledge is a matter of tracking the truth and I have argued that the modal dimension of the notion of epistemic control consists in an inquiring agent's doxastic states changing to suit varying hypothetical situations. I have also explored Nozick's more specific idea that the range of possible scenarios in which the inquiring agent should succeed (or not fail) in order for her actual belief to qualify for knowledge can be represented using subjunctive conditionals. I have evaluated the pros and cons of three of these conditions: section 5.3 has been devoted to the sensitivity condition, section 5.4 to the receptivity condition and section 5.5 to the safety condition. The conclusion has been that the notion of epistemic control cannot be specified in terms of sensitivity or receptivity because they are not necessary for knowledge. The safety condition, by contrast, is suitable to that task. Not only because it is necessary for knowledge, but because, as I have explained in section 5.6, there is a more general connection between the notions of control and safety in general. What we need now is an argument for the hypothesis that knowledge requires epistemic control. In particular, we need a conceptual connection between the notion of control and the concept of knowledge. As we will see in the next chapter, the connection will be provided by the idea that knowledge is a cognitive achievement.

6

KNOWLEDGE AS ACHIEVEMENT AND THE CONTROL THEORY OF KNOWLEDGE

This chapter starts with the basics: a presentation of Ernest Sosa's initial motivations for the virtue-theoretic project in epistemology until its most recent development as a *performance-based epistemology* (sections 6.1 and 6.2). Particular emphasis is put on the question of what makes virtue epistemology a distinctive approach to epistemic normativity. The main focus of the chapter is on the so-called *achievement theory of knowledge*,[1] a leading approach based on Sosa's performance-assessment model that conceives knowledge as a *cognitive achievement*. Section 6.3 introduces the standard conception of achievement (the view that achievements are successes because of ability) and the standard achievement theory of knowledge, according to which knowledge is a cognitive achievement understood as a cognitive success because of cognitive ability (viz., as belief that is true because of cognitive ability). Section 6.4 proposes an original alternative to the standard conception of achievement: to define achievements in terms of control over performance. Section 6.5 uses a recent objection made by Duncan Pritchard against the standard achievement account of knowledge (the problem of lucky achievements) to motivate an alternative version of the achievement theory of knowledge, a theory that I will call (in section 6.6) the *control theory of knowledge*. Like the standard achievement account, the control theory of knowledge defines knowledge as a cognitive achievement. Unlike the standard version, it does not understand the notion of achievement merely as success *because of* ability, but as success over which the agent has *control*. In this way, the account defines knowledge in terms of *epistemic control*, a notion that will be fully defined only at the end of chapter 9. Section 6.7 addresses another objection made by Pritchard against the standard achievement theory of knowledge (the problem of easy achievements). The discussion of this problem will help to refine the control theory of knowledge.

6.1 THE GENESIS OF VIRTUE EPISTEMOLOGY

Ernest Sosa introduced the notion of *intellectual* or *epistemic virtue* in his classic paper "The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge" (1980) as a way of overcoming the incompatibility between coherentist and foundationalist conceptions of the structure of knowledge. *Coherentism* and *foundationalism*

1 I borrow the terminology from Turri (*forthcoming*).

were characterized by means of two accurate metaphors about the structure of a given mind's knowledge system. The foundationalist structure of knowledge is a *pyramid*, Sosa explains, in the sense that known beliefs that stand at the apex are being supported by known beliefs at the lower levels, which are ultimately justified by a foundation of non-inferential knowledge at the base. At the base, one might find *cogito* beliefs (e.g., the belief that I am thinking), but also basic beliefs coming from perception or memory (the configuration of the knowledge base depends on the version of foundationalism held). By contrast, the coherentist structure of knowledge is a *raft*, in the sense that the different parts of the structure are horizontally tied to each other without any of them enjoying a fundamental status. The 'glue' between the different parts of the raft (i.e., between the agent's beliefs) is the coherence relation, which may be conceived in several distinct ways.[2]

How does Sosa bring together coherentism and foundationalism?[3] Sosa's strategy consists in arguing that both views are instances of a more general view that he calls *formal foundationalism*, the view that epistemic properties such as knowledge or justification, which are *normative* or *evaluative*, supervene on specifiable *non-evaluative* base properties.[4] This view is grounded on the widely-held thesis that evaluative properties (such as right, wrong, good, bad or worthy) supervene on non-evaluative properties, i.e., properties that can be *neutrally described*.[5] Which would be then, according to foundationalism and coherentism, the relevant non-evaluative base properties upon which knowledge and justification supervene? Examples of foundationalist base properties are the property of being grounded on *cogito* beliefs or the property of being grounded on perception. As per coherentism, the relevant base property is coherence among a set of beliefs.[6]

It was Sosa's next step that led to the genesis of *virtue* epistemology: he generalized the normative framework of virtue ethics and applied it to epistemology. The core idea of virtue ethics is, roughly, that

---

2 See chapter 7, fn. 13.

3 The following reconstruction of Sosa's arguments for the integration of foundationalism and coherentism into a virtue-theoretic framework is in large part based on Turri's excellent presentation of Sosa's bi-level virtue epistemology (Turri 2013).

4 That *A* supervenes on *B* roughly means that there is no difference in *A* without a difference in *B*.

5 Descriptions such as "This is a watch" are neutral; descriptions such as "This is a good watch" are not.

6 One might think that coherence when understood in logical terms (e.g. as logical consistency) is a normative concept because logic is normative. Although logic might be normative, for instance, for thought in so far as it tell us how to reason, the kind of normativity that is at issue here is different. One thing is that *X* is normative because *X* is some rule, model, standard or ideal to which *Y* ought to conform (or, more strongly, which *Y* ought to follow). Another thing is that *X* is normative because it can be neutrally described, unlike other concepts that, intuitively, cannot. In this latter sense, logical consistency is not a normative property, whereas properties such as being good, worthy or wrong are.

morally right actions are actions that stem from stable moral virtues. For example, Sosa asks: why is a doctor justified in assisting the birth of baby Hitler? Sosa answers:

> His action is the result of certain stable virtues, and there are no equally virtuous alternate *dispositions* that, given his cognitive limitations, he might have embodied with equal or better total consequences, and that would have led him to infanticide in the circumstances. (Sosa 1980: 23)

When Hitler was born, nobody could imagine the atrocities that he would commit when adult. Had refused the doctor to assist Hitler's birth or had he killed the baby at that time, he would have been blamed for it, because in those hypothetical cases the doctor would have acted against his character, and namely against his disposition to heal his patients. Of course, things would have been very different if the doctor had possessed the power of predicting the future, but that would have required a different set of dispositions that he did not possess at the time (hence Sosa's qualification 'given his cognitive limitations').

The virtue-theoretic approach to ethics does not regard acts, consequences or intentions as the primary bearers of moral qualities. Instead, primary bearers of moral qualities are the *agent's dispositions*. It is this aspect of the theory that was of particular interest to Sosa, since to have a disposition to φ is a *non-evaluative basis* upon which the *normative* property of being morally justified supervenes. If we apply this piece of theory to the epistemic case it turns out that to have a *disposition to form true rather than false beliefs* is a *non-evaluative basis* upon which the *normative* property of being *epistemically* justified supervenes. This was Sosa's master move. In this way, he conceived *intellectual* or *epistemic virtues* as dispositions to believe truths (rather than falsehoods) to which epistemic justification primarily attaches. On this view, our perceptual abilities count as epistemic virtues, because they normally bring us to the truth.

Having a disposition to form true rather than false beliefs is not the only criterion that is necessary to make an agent's cognitive disposition an intellectual virtue: it must also be integrated with the rest of the agent's cognitive dispositions (a device implanted in one's brain might dispose one to form true beliefs but if that device operates independently of one's agency, then it does not count as an intellectual *virtue*). We can give the following definition of epistemic virtue:

- Epistemic Virtue: *S*'s cognitive disposition to form beliefs counts as an epistemic virtue of *S* if and only if (i) it is globally reliable, in the sense that it tends to produce true rather than false beliefs about propositions of certain field (e.g., propositions about colors) under certain types of circumstances (e.g., circumstances

> with good light conditions), and (ii) *S*'s cognitive disposition is sufficiently integrated with other of *S*'s cognitive dispositions.[7]

How are beliefs justified according to the virtue-theoretic approach? Sosa's idea is that in the same way as primary justification attaches to moral virtues rather than to its acts, primary justification attaches to epistemic virtues rather to beliefs, which are *derivatively* or *secondarily* justified because of being the outputs of epistemic virtues.[8] For example, perceptual beliefs are justified simply in virtue of being the output of reliable perceptual abilities.

Where do foundationalism and coherentism stand in this picture? Both views may be interpreted in virtue-theoretic terms. Is it intellectually virtuous to believe *p* when *p* coheres with the rest of believed propositions? It is, because the property of being coherent contributes towards getting one to the truth. What about foundationalism? Is it intellectually virtuous to believe *p* when *p* is grounded on, say, perception? It is, because forming beliefs on those grounds gets one to the truth. This is how Sosa brought together two accounts of knowledge and justification that were considered incompatible at the time. But more importantly, this is how virtue epistemology was born as a new distinctive approach to *epistemic normativity*.

6.2 PERFORMANCE-BASED EPISTEMOLOGY

In the next two decades, Sosa and other epistemologists applied the virtue-theoretic framework to important issues in epistemology such as the nature of knowledge or justification, skepticism,[9] the Gettier problem,[10] the *Meno* problem[11] or the New Evil Demon problem.[12] I will not review here the merits and shortcomings of virtue epistemology (although I will be concerned with a central question it tackles: the question of what is knowledge). For the moment, my main focus will be rather on epistemic normativity.

---

7 Two remarks. First, there are no evaluative terms in this definition, so that it provides a proper non-evaluative basis upon which justification and knowledge, which are normative concepts, may supervene. Second, a terminological remark: virtue epistemologists use indistinctively the terms 'virtue', 'competence', 'ability', 'disposition', 'power', 'capacity' and 'faculty'. As Turri (2013) points out, these are mere verbal variations that should not be taken to indicate a shift in the underlying views. I will also use the terms interchangeably, unless otherwise indicated.

8 This phenomenon is what Sosa calls the *stratification of justification*.

9 Mainly to argue against the Cartesian skeptic.

10 The problem of elucidating why Gettier subjects do not know.

11 The problem of elucidating why knowledge is more valuable than mere true belief.

12 A problem for externalism. Internalist argue that in demon worlds in which one is systematically deceived (e.g., perceptually deceived) one is still justified in one's beliefs (because some reflective access requirement is met) despite the fact that one does not form them via reliable belief-forming processes. The argument aims to show that a reliability requirement is not necessary for epistemic justification. See chapter 4, fn. 20.

In his two-volume book (2007/2009), Sosa expands the normative framework of virtue epistemology by presenting it as an instance of the broader framework of performance normativity. This turns virtue epistemology into a *performance-based epistemology.* Sosa's general idea consists in conceptualizing belief formation as a kind of cognitive performance so that the normative properties of performances in general may be used to account for the concept of knowledge.

6.2.1 *Performances and Normative Assessment*

When we see other people act we typically evaluate their performances. For example, we make moral evaluations of other people's acts. However, the evaluative properties that are relevant to the question of what is knowledge are obviously not the properties of the moral domain but the kind of properties we ascribe to a performance in terms of whether it attains its aim, i.e., the kind of properties we ascribe to a performance *qua* performance.

First, all performances have an aim, i.e., a course of action (someone or something behaving in certain way) does not intuitively count as a performance unless some aim is pursued. A stone falling to the ground, for example, does not count as a performance of the stone because stones have no aims. By contrast, a heart that pumps blood is said to perform because it is its biological aim to do so. Therefore, all performances have an aim.

Here is a possible counterexample to this claim: someone's fall might be considered a performance even though this person had no intention to fall. Here is a possible reply: if that person's fall is said to be a performance, it is because that person was engaged in another (perhaps broader) performance (e.g., walking to work or strolling by pure pleasure). That is, since the person's fall contributes to the failure of the performance in which she was already engaged, we implicitly consider her fall part of that broader performance and so we claim that her fall is a performance. Plausibly, all performances have an aim.

Second, as Sosa points out a performance may be *consciously* aimed at certain outcome (e.g., PhD students aim at writing dissertations), but not necessarily (e.g., a heartbeat non-consciously aims to make the blood to circulate) (Sosa 2011: 14).

Third, Sosa calls *endeavors* performances that are, so to speak, focused on an aim. His general idea is that "[a]n endeavor (...) has its essential aim, an aim inherent in it" (*ibid.*). As far as I can see, this may be for two reasons: a performance counts as an endeavor either because its aim is recognized as the sort of aim that performances of that type typically have (e.g., pumping blood is supposed to be the aim of a heart) or because an agent attempts (not necessarily consciously) to attain that aim with a significant degree of effort (in so doing the agent makes that aim his own, in a very strong sense).

Especially in the case of endeavors but also in the case of other performances, the fact that they are goal-directed makes them objects of normative assessment. In particular, it can be assessed whether an endeavor or a performance attains its aim (succeeds) and whether it does it an correct manner. It is important to emphasize that a consequence or an outcome of an endeavor or a of performance counts as a success only if the consequence or the outcome is something at which the endeavor or the performance aimed. Sosa illustrates the point with an example of a performance with 'nested' aims:

> [O]ne can aim to flip a switch by operating on it in a certain way with one's fingers. In one's plan the switch-flipping endeavor might itself serve a further endeavor: one might be aiming to turn on the light. One might also do something else thereby, such as alerting a prowler, even if this last is not assessable for its degree of success. One's alerting of the prowler is not a "success" *if it was not one's aim*. (Sosa 2011: 14; emphasis mine)

In part I, we saw that we ascribe control to something or someone only if it or she has an aim. This is why we ascribe control to hearts, humans and other beings but not to stones. Might be the consequences of one's performance under one's control if they are not something at which one aimed by performing as one did? For example, would be one's alerting of the prowler something over which one has control if one did not aim at attaining that outcome? The answer (to both questions) is negative. If by turning on the light one does not aim to alert the prowler, one's alerting of the prowler is by luck and hence not under one's control. In sum, in the same way as certain outcome does not count as a success if one does not aim at it by performing as one does, one does not have control over the outcome of one's performance if one's performance does not have as aim to attain such an outcome (no matter that the aim is implicit or unconscious).

6.2.2 *Sosa's Performance-Assessment AAA Structure*

According to Sosa (2007: 22-3), performances may be assessed along three normative dimensions:

1. *Accuracy*: Is the performance successful, i.e., does the performance attain its aim?

2. *Adroitness*: Is the performance the product of competence or ability?

3. *Aptness*: Is the performance successful *because of* competence or ability, i.e., accurate because of adroit?

An archer's shot is *accurate* when the arrow hits the target; *adroit* when the archer is skillful and *apt* when the arrow hits the target due to or because of the archer's skills.

6.3 THE ACHIEVEMENT THEORY OF KNOWLEDGE

Sosa's account of knowledge is an application of the performance-assessment AAA structure to belief formation, which he conceives as a performance, a cognitive one though.

6.3.1 *The Aim of Belief Formation*

What could be the aim of belief formation? One plausible answer is that belief formation is biologically aimed at truth. Recall Nozick's evolutionary explanation of how we came to have truth-tracking capabilities. The explanation began with the plausible assumption that possessing the biological preconditions of belief is an adaptive advantage connected with the advantages of having true beliefs, and that the evolutionary process can give organisms true beliefs only by giving them belief-forming mechanisms with a *tendency* to form true rather than false beliefs (Nozick tried to further justify that such a tendency is grounded on a capability to track changes in variable situations, i.e., a truth-tracking mechanism).[13]

The evolutionary explanation fits in well here: since our innate cognitive abilities (e.g., our perceptual capacities) have an evolutionary-shaped tendency to form true rather than false beliefs, belief formation is biologically aimed at truth. Although this seems true, Sosa finds plausible the possibility of forming beliefs without primarily aiming at truth, but at some pragmatic goal:

> There is such a thing as wishful thinking of a sort that aims at the intellectual comfort of the believer. For example, we are said to systematically overestimate our own beliefs. Such beliefs can aim at our comfort *regardless of truth*, which in some cases might not even be *an* aim, much less *the* aim. (Sosa 2011: 15)

According to Sosa, the only candidates for knowledge are beliefs that result from cognitive performances that have truth as the primary aim, i.e., cognitive performances in the *endeavor* to attain the truth.

13 Other (implausible) explanations of such a tendency to believe the truth were that 1) there is a pre-established big set of true beliefs that we have assigned from birth and that 2) some super-being constantly intervenes to prevent us from believing falsehoods.

However, since cases like wishful thinking are not that unusual, we may ask: what degree of commitment with the truth is required in order for an endeavor to be assessed under the AAA structure? Is a belief evaluable as true just because it is true? Is it perhaps needed a more substantial (albeit still weak) commitment to the truth, such as the commitment we have in virtue of the automatic aiming at truth of our evolutionary-shaped belief-forming mechanisms? Or is it required a very substantial commitment, such a sincere *desire for truth* or the possession of some *character trait* that involves a high motivational profile (e.g., *conscientiousness*)? The *responsibilist* branch of virtue epistemology (e.g., Montmarquet 1993, Zagzebski 1996) will surely embrace the latter option.[14]

Sosa (2011) seems to take the middle way. He neither thinks that a true belief is evaluable as true just because it is true, nor he would consider sufficient the weak commitment to the truth that we have in virtue of our evolutionary-shaped belief-forming abilities. For Sosa, a more substantial commitment is needed, one that guarantees that the relevant cognitive performances have truth as *the* aim. However, he does not think that such a commitment needs be the output of character traits such as conscientiousness (a trait that has a marked motivational profile) or of a deliberative, voluntary or conscious process. The relevant commitment may operate as prejudices do: in an unconscious way. For instance, think of an agent (e.g., a racist) with some subconscious bias that automatically makes her disregard known evidence coming from science in matters concerning race. In that case, the racist may be said to be in an endeavor, but not to attain truth. The kind of commitment to the truth that Sosa has in mind plausibly operates in the same way, but in the opposite direction, i.e., towards truth.[15]

6.3.2 *Knowledge as a Cognitive Achievement*

According to Sosa, only cognitive performances in the endeavor to attain truth (in the sense sketched above) fall under the AAA structure:

1. *Accuracy*: Is the belief true?
2. *Adroitness*: Is the belief the product of reliable cognitive ability?
3. *Aptness*: Is the belief true *because of* the exercise of reliable cognitive ability?

14 Unlike virtue reliabilists, virtue responsibilists place a great emphasis on epistemic responsibility. Accordingly, rather than conceiving epistemic virtues as reliable dispositions to believe the truth (such as our perceptual faculties), they take virtues to be *character traits*, such as conscientiousness or open-mindedness.

15 In cases of reflective knowledge, there might be an ascent to consciousness, as when someone makes explicit one's prejudices and one reflects about it.

In Sosa's view, when the answer to the third question is positive, i.e., when a belief is apt, the belief qualifies as knowledge. In recent years, this way of thinking about knowledge has lead to an increasingly popular conception of knowledge as a cognitive achievement. Call the view that holds that knowledge is a cognitive achievement the *achievement theory of knowledge*:

- ACHIEVEMENT THEORY OF KNOWLEDGE (ATK): Knowledge is a cognitive achievement.

What is an achievement, in general, and a cognitive achievement, in particular? Depending on how the notion of achievement is conceived, ATK will be specified differently.

6.4 ACHIEVEMENT AS SUCCESS BECAUSE OF ABILITY

The shot of an archer is said to be her achievement when the arrow hits the target because of the exercise of the archer's abilities. This leads to what we may call the *standard conception of achievement*:

- STANDARD CONCEPTION OF ACHIEVEMENT: The success of *S*'s performance is an achievement of *S* if and only if *S*'s performance is successful because of the exercise of *S*'s abilities.

As Turri (*forthcoming*) points out, success or achievement talk here is not intended to imply any sort of approval. Consider the following example. God's making the waters flood the Earth thus committing a genocide might be considered an *achievement* of God simply because his action attains its goal (to punish humanity) and regardless of the fact that most people would be highly disappointed and would not give moral approval. Recall that the relevant dimension of assessment that is relevant here is that of performances *qua* performances, not the moral dimension.

6.4.1 *The Standard Achievement Theory of Knowledge*

By applying the standard conception of achievement to ATK, we come up with what we may call the *standard achievement theory of knowledge*:

- STANDARD ACHIEVEMENT THEORY OF KNOWLEDGE (SATK): Knowledge is a cognitive achievement, where a cognitive achievement is a belief that is true because of cognitive ability.

In recent years, SATK has been defended or endorsed by several important epistemologists (Greco 2010; Sosa 2007, 2011; Riggs 2002, 2009; Turri (*forthcoming*) and Zagzebski 2003). However, some formidable objections has been raised against it. In particular, Duncan Pritchard

(2008, 2009b, 2010a, *et al*. 2010) has argued that SATK faces two serious problems.

On the one hand, according to Pritchard some successes that are because of ability (i.e., some apt successes) may be by luck, e.g., the competent shot of a skillful archer is by luck if it hits the only target among thousands of possible targets that is not protected with an invisible force shield that would have repelled any incoming arrow. Pritchard argues that if such successes count as achievements, then we seem compelled to treat cognitive successes that are relevantly analogous as also being achievements, which would mean that we would thereby be compelled to regard the cognitive achievements in question as knowledge, even despite the luck involved (Pritchard *et al.* 2010: 35). He exemplifies the point with cases in which an agent identifies an object (e.g., a barn) but in which the agent could easily have misidentified replicas of the object (e.g., fake barns). Pritchard calls this the *problem of lucky achievements*.

On the other hand, according to Pritchard some successes because of ability (i.e., some apt successes) do not count as achievements because of the ease with which they are brought about, e.g., the raising of one's hand or blinking. In the same way, basic perceptual true beliefs do not count as cognitive achievements because of the ease with which they are brought about. Pritchard calls this the *problem of easy achievements*. In what follows, I will analyze these problems in detail and I will sketch a particular version of ATK that steers clear of them.

6.5 ACHIEVEMENT AS CONTROL

I propose a different conception of the notion of achievement to the one assumed by defenders of SATK. I call it the *control conception of achievement*:

- CONTROL CONCEPTION OF ACHIEVEMENT: The success of *S*'s performance is an achievement of *S* if and only if *S* has control over her performance (alternatively: just in case *S* has control over the success).[16]

Consider the following example:

> ARTEMIS I
>
> Artemis is a very skillful archer. She takes aim and shoots. In normal conditions, the shot would have hit the bull's-eye. The wind, however, is abnormally strong, and just strong enough to divert the arrow so that, in conditions thereafter normal, it would miss the target altogether. However, shifting winds guide it gently to the bull's-eye after all. (Adapted from Sosa 2007: 22)

16 The term control in this definition is to be understood in the ordinary sense of the term.

Ordinarily, we would not say that the shot (its success) is under Artemis's control. Therefore, according to the control conception, Artemis's success is not an achievement because she lacks control over the shot.[17]

§ *Refining the control conception of achievement*. What kind of control does Artemis lack, tracking and effective control, just effective control or just tracking control? Recall the definitions of tracking and effective control offered in chapter 3:

- TRACKING CONTROL: *A* has tracking control over *B* if and only if *A* monitors *B*.
- EFFECTIVE CONTROL: *A* has effective control over *B* if and only if (i) it is *A*'s aim that *B* is in certain state *S* (ii) *A* has a disposition to cause/determine *B* to be in *S*, (iii) *B* is in *S* and (iv) *B* is in *S* because of *A*'s disposition to cause/determine *B* to be in *S*.

Does Artemis lack tracking control, i.e., does she fail to monitor the relevant parameters that are salient in the context? Although she is an expert archer and she can monitor the wind, light and distance conditions reliably (as well as the state of the bow and the arrows, her breathing, and so on), this time she fails to monitor the gust of wind that ultimately makes her arrow hit the target. Unexpected gusts of wind might be difficult to monitor, of course, but the level of monitoring competence required might vary from situation to situation. It seems that Artemis cannot be properly said to have tracking control, because she does not achieve the degree of monitoring required to true ascription of tracking control in that situation (presumably, detecting unusual gusts of wind is very difficult).

Does she also lack effective control? She clearly lacks effective control because, although (i) it is Artemis's aim to make the arrow hit the target, (ii) she has the ability to do it and (iii) the arrow actually hits the target, the fact that (iv) the arrow hits the target is *not* because of the exercise of her archery skills, but to the double intervention of the wind.

The control conception of achievement says that a success is an achievement just in case the agent has control over her performance, where the relevant sense of the term 'control' is the ordinary one. In chapter 3, I explained that in ordinary discourse we say that *A* controls *B* either when (i) *S* has effective control over *B*, when (ii) *A* has tracking control over *B*, or when (iii) *A* has both effective and tracking control over *B*. The third option is the most common one, but there are plausible examples of (i) and (ii) because, depending on the context, tracking or effective control might be the only forms of

17 According to the standard conception of achievement, Artemis's success is not an achievement either, because the fact that the arrow hits the target is not because of her archery skills.

control that are relevant to evaluate ascriptions of the form '*A* controls *B*'.

By way of illustration, consider the statement 'The doctor controls the patient's infection'. The statement is true in a context in which the doctor gives some specific antibiotics to the patient in order to combat the infection (the salient sense of control is effective control). But it is also true in a context in which the doctor runs some tests to determine what causes the infection (the salient sense of control is tracking control). Finally, the statement is true in a context in which the doctor runs tests and administers antibiotics (the salient senses of control are both tracking and effective control). In all these contexts, we can ordinarily and truly say of the doctor that she controls the patient's infection.

Now, if we explain the notion of achievement in terms of control over performance, we should expect our attributions of achievement to have the same truth-conditions as our attributions of control. That is, my contention is that if in an context one can truly say from an ordinary point of view that *S* has control over her performance, which results in outcome *O*, then we can also truly say that *O* is an achievement of *S*, and *vice versa*, if in a context one can truly say that *O* is an achievement of *S*, then one can truly say from an ordinary point of view that *S* has control over her performance, which results in *O*.

Given the aforementioned ambiguity of the ordinary notion of control, we can expect contexts in which we truly say that *O* is an achievement of *S* because *S* has effective control, contexts in which we truly say that *O* is an achievement of *S* because *S* has tracking control, and contexts in which we truly say that *O* is an achievement of *S* because *S* has both effective and tracking control. Accordingly, we can give a more accurate definition of achievement by disambiguating the use of the term 'control' in the definition given above:

- Control Conception of Achievement (revised): The success of *S*'s performance is an achievement of *S* if and only if *S* has control over her performance, where *S* has control over her performance if:
  - *either* (i) *S* has effective control over her performance,
  - *or* (ii) *S* has tracking control over her performance,
  - *or* (iii) *S* has both tracking and effective control over her performance.

Which sense or senses of control are relevant in each case depends on which type of control is salient in the context (recall the examples of the doctor and the patient's infection).[18] A caveat is in order: in some

18 An alternative approach to the notion of achievement (also in terms of control) is the following: the kind of control that is relevant in order for a success to be an achievement depends on the goal of the performance that is being assessed. The idea,

contexts we find acceptable to say that *S* has control over her performance when, let us say, *S* has tracking control, but without that necessarily implying that *S* lacks effective control over her performance. The point is that, in some contexts, the possession of effective control is irrelevant to assess the statement "*S* has control over her performance" for truth, because tracking control is the only sense of control that matters in the context. As we will see next, the disjunctive nature of the control conception of achievement steers clear of the problem of lucky achievements, which according to Duncan Pritchard (2008, 2009b, 2010a, *et al*. 2010) proves the standard conception of achievement wrong.

6.6 THE PROBLEM OF LUCKY ACHIEVEMENTS

We can split the problem of lucky achievements into two parts.

§ *Part one*. According to Pritchard, some successes because of ability may be by luck, i.e., some achievements (conceived in the standard way) may be by luck. He gives the following example:

> ARTEMIS II
>
> Artemis takes aim. Everything is normal: soft winds, usual distance. She shots and hits the bull's eye. However, she has randomly shot at the only target among thousand of possible targets that is not protected with an invisible force shield that would have repelled any incoming arrow. (Adapted from Pritchard *et al*. 2010: 35)[19]

Artemis's success is *due to* the exercise of her shooting skills, so that, according to STANDARD CONCEPTION OF ACHIEVEMENT, her successful shot counts as an achievement. The problem is that the environment is such that she could easily have missed the shot, which means that she was at risk of missing and hence that her success (and therefore her achievement) is by luck.

§ *Part two*. Pritchard thinks that if Artemis's success is an achievement (according to the standard conception of achievement it is) and if an achievement like Artemis's is by luck (intuitively, it is) and if knowledge is a cognitive achievement (according to SATK, it is), then *cognitive achievements* that are relevantly analogous to Artemis's may be by luck. However, knowledge is incompatible with luck. Therefore, knowledge is not a cognitive achievement (contrary to what SATK says).

---

then, is that the context does not make salient the relevant type of control but the kind of goal that is relevant to assess the performance in question and, derivatively, the kind of control needed to accomplish that goal.

19 I already used this case in chapter 4 when discussing environmental veritic luck.

My position with respect to the problem of lucky achievements is the following: I agree with Pritchard that some achievements are by luck (part one) but I disagree that cognitive achievements that are relevantly analogous to Artemis's are by luck (part two). In addition, while I think that ATK is true (knowledge is a cognitive achievement), I do not uphold SATK (knowledge is a cognitive achievement, where a cognitive achievement is understood as a cognitive success because of cognitive ability), because I reject the standard conception of achievement (the one assumed by SATK). Let us proceed part by part.

6.6.1 *Reply to the Problem (part I)*

As I explained before, once we explain the notion of achievement in terms of control over performance, we should expect our attributions of achievement to have the same truth-conditions as our attributions of control. Is Artemis's shot in Artemis II an achievement? Intuitively, it is (independently of one's conception of achievement). Pritchard shares this intuition. After all, he claims, Artemis's shot is accurate because of the exercise of ability (he has the standard conception of achievement in mind, though).

Would we also say that Artemis's shot is under her control from an ordinary point of view (and, therefore, would it follow from the control conception that Artemis's shot is an achievement)? I think we would say that. Artemis, after all, shoots in a very proficient manner, so quite plausibly we would say from an ordinary point of view that she has control over her shot. To strengthen the intuition suppose that the shooting distance is 180 meters and that she is the only archer in the whole world capable of shooting at such a long range. The question is: in virtue of what type of control is that ordinary claim true?

Artemis, in shooting as she does, completely fails to monitor the force shields of the targets. Consequently, she does not have tracking control over her shot. Note, however, that I do not deny that she successfully carries out some sort of monitoring that is relevant to evaluate whether she has tracking control or not. She does. For example, she successfully monitors the strength and direction of the wind and the state of the bow and the arrows. However, the degree and sort of monitoring exhibited by Artemis does not suffice for true attribution of tracking control *in those circumstances*. For any circumstances *C*, the degree and the kind of monitoring required to have tracking control in *C* is set by pragmatic factors of *C*.

In Artemis II, the presence of *invisible* force shields makes the kind of monitoring required for having tracking control a very demanding form of monitoring given Artemis's archery capabilities (or given the capabilities of any other archer, as archery ranges with targets protected with invisible force shields are not the kind of ranges to which regular archers are used). To carry out the kind of monitoring required one would need to use some detection device or to possess some superpower. Artemis lacks both and therefore lacks tracking control. Therefore, if the statement 'Artemis has control over her shot' is true in the context of ARTEMIS II (it certainly seems true), it is in virtue of the fact that Artemis exhibits a great degree of effective control: she is the best at long distance shots and her shot is successful *because of* her shooting abilities.

Effective control over a performance, however, might not be sufficient to prevent the success that results from the performance from being lucky. In ARTEMIS II, it is undeniable that Artemis's success is by luck. Pritchard's intuition is that there is also achievement (he has in mind the standard conception of achievement, i.e., achievement as success because of ability). The control conception of achievement captures all the intuitions: there is control (of the effective sort), achievement, and luck.

6.6.2 *Partial and Complete Achievements*

According to CONTROL CONCEPTION OF ACHIEVEMENT, attributions of achievement have the same truth-conditions as attributions of control. Accordingly, there should be cases in which, for example, an agent has effective control, lacks tracking control and we do not consider her success an achievement because we do not consider that she has control over her performance in virtue of her manifest lack of tracking control. These would be cases structurally similar to ARTEMIS II but in which the only kind of control that is relevant for true ascription of control *in that context* is tracking control. Consider the following case:

> ARTEMIS III
>
> Artemis, the skillful archer, participates in an archery competition in which failing to detect the invisible force shields of the targets automatically disqualifies the participant. In fact, finding the only target without a force shield before the other participants is the most important thing in the competition: if one shoots at it, one wins, *regardless* of whether one ends up hitting the target. That is, once one detects the only target without of a force shield (among, say, hundred of possible targets), one just has to shoot at it to win (no matter whether the shot is successful). However, this is something that Artemis ignores. She *randomly* selects a target, takes aim, shoots and hits the bull's eye. Luckily, the target she has picked is the only one without a force shield. She is disqualified.

In this competition, target selection is the only task in which one must be competent. Artemis success in selecting the right target is anything but competent, because it does not manifest any capability to monitor the parameters that the competition establishes. Her success in picking the target is thus due to sheer luck. As in ARTEMIS II, I do not deny that Artemis successfully carries out some sort of monitoring (it is undeniable that she monitors the wind, the bow, the arrows, the light and distance conditions, and so on). However, the rules of the competition make salient that one must detect the force shields. The point is that failing to carry out that task implies failing to achieve the degree (and kind of) monitoring required to competitors for having tracking control over their shots.

Since Artemis lacks tracking control over her shot and having effective control is not important in the context, the control conception of achievement implies that Artemis's successful shot is *not* an achievement (i.e., the ordinary ascription 'Artemis controls the shot' is not true in the circumstances because she does not have the relevant sense off control picked out by the context). To put in other terms, pragmatic factors of the circumstances make salient one of the disjuncts of CONTROL CONCEPTION OF ACHIEVEMENT; since that disjunct is not satisfied, the ascription 'Artemis controls her shot' is false in the context and her success is not an achievement. Consider now this other case:

> ARTEMIS IV
>
> Artemis, the skillful archer, participates in an archery competition in which shots are assessed along two dimensions: judges give one to ten points for target selection and one to ten points for accuracy (ten points means excellent performance). The sum of the points obtained on each task is the final score. The winner of the competition is the one with a higher final score. Artemis ignores that all the targets except one are protected with invisible force shields. She quickly selects one at random (one which is at a great distance), takes aim, shoots with great mastery and hits the bull's eye. Luckily, the target she has picked is the only target without a force shield. The judges give her ten points for hitting the bull's eye and one point for selecting the target. She does not win the competition.

In this case, we have a broader perspective of assessment: the type of control that matters is both tracking and effective control. From this broader perspective, one feels that Artemis's shot is not as bad as in ARTEMIS III, but still it does not count as an achievement, as one has the impression that, although it is a great shot from a great distance, there is something amiss with the way she picks her target, namely with the fact that she does it randomly, and this is something that one takes into consideration when it comes to judging whether the shot counts as an achievement or not. After all, the judges only give to her eleven points for her performance, a rather good score but definitely not an exceptional one. The point is that Artemis's successful shot is not a *complete* achievement, in the sense that it might be disputed or questioned from some dimension of assessment (namely from the target selection dimension).

The control conception of achievement correctly predicts this sort of evaluations: since condition (iii) of CONTROL CONCEPTION OF ACHIEVEMENT is only partially satisfied (Artemis has only effective control over the shot), her success counts as a partial achievement, but not as a complete one. Call a *complete achievement* an achievement that is indisputable or unquestionable from any dimension of assessment.[20] What makes us dispute the achievement of an agent? For the most part, the presence of luck in the way the agent succeeds. Moreover, an achievement that involves luck in the way the agent succeeds might be disputed even though in the context in which the agent performs the presence of luck does not prevent her successful performance from being an achievement. Let me exemplify the point.

In ARTEMIS II, Artemis's shot is successful and is considered an achievement: the presence of luck, as we have seen, does not prevent the shot from counting as an achievement because Artemis has effective control over the shot. However, it is by luck that her shot is successful (Artemis lacks tracking control over the shot). This makes us question or dispute the shot. Consider now a winner of the competition described in ARTEMIS III. Suppose that, besides selecting the target in a very proficient manner, the archer hits the bull's eye because of the intervention of unexpected shifting winds (here the archer would have tracking control but would lack effective control over the shot). Although the shot counts as an achievement from the relevant context of assessment (i.e., from the point of view of the competition), the shot might be questioned from a broader perspective: one that takes into account that the shot hits the bull's eye by luck. From this broader perspective, the relevant achievements in these two cases are not complete but partial, because an achievement that is compatible with luck can always be disputed. The idea can be put in terms of control:

- COMPLETE ACHIEVEMENT: *S*'s success is a complete achievement of *S* if and only if *S* has effective and tracking control over her performance.

In view of this, we can now reinterpret the problem of lucky achievements: it only shows that achievements that are not complete might

20 Keep in mind that the dimension of assessment that is relevant here is whether the relevant performance attains its aim, i.e., the dimension of assessment of the performance *qua* performance. In this way, a complete achievement might be disputed from a moral point of view and still count as complete.

be by luck. Complete achievements are not only incompatible with luck, but also with fortune. To illustrate the point, consider the following case:

> MAGNETIC ARROW
>
> Artemis takes aim, shots and hits the bull's eye. She has randomly shot at the only target among thousand of possible targets that is not protected with an invisible force shield that would have repelled any incoming arrow. However, had she shot at another target, she would have nevertheless hit the same target that she has actually hit. The reason is that, unbeknownst to Artemis, 1) the target she has picked is the only metallic target in the area, 2) her arrow has a strong electromagnet at the tip and 3) if she had shot at another target, a hidden agent would have activated the electromagnet.[21]

MAGNETIC ARROW is a case of fortune because there is no risk that Artemis fails to hit the target.[22] Her achievement (hitting the bull's eye with great proficiency) is not complete, because she fails to monitor important factors of the situation (the force shields, the hidden agent, the electromagnet, and so on). Again, the sort of monitoring required to have tracking control over the shot is very demanding, but so are the circumstances.

The point, in any case, is that if to achieve a goal *completely* is to perform with tracking and effective control, then no complete achievement can be by luck or by fortune: as we have defined the notions of luck and fortune, no performance can be successful by luck or fortune if the agent has control over it.

§ *An objection considered*. Think about again the archer who wins the competition described in ARTEMIS III because she selects the only target without a force shield in a very proficient manner but whose shot is successful by luck. One could argue: why should we consider her achievement partial? After all, she wins the competition and consequently her achievement seems a complete one. True, it is a complete one, but only in local sense. Shooting performance, in general, involves two aims: selecting the target well and hitting the target. The winner of the competition successfully selects the target and successfully hits it. Concerning the first aim, her achievement is impeccable; she selects the target proficiently, in a way that does not involve luck. The context of the competition makes salient this aim and in this sense the shot of the winner is a complete achievement, because she has the relevant sense of control that is required *in the context*. We can

21 The example is inspired by a similar case by Gundersen (2010).

22 Recall that the difference between luck and fortune has to do with the risk that the event in question (in this case Artemis's successful shot) had of not occurring.

call achievements of this kind *locally complete achievements*. Here is a definition:

- LOCALLY COMPLETE ACHIEVEMENT: Given a context of performance *C*, such that in *C* the salient aim of *S*'s performance is *A*, *S*'s success in achieving *A* is a locally complete achievement of *S* if and only if *S* has the form of control (effective, tracking or both) that is required by *C* for achieving *A*.

In the context of the competition described in ARTEMIS III, the salient aim of the competitors is to select their targets well, and the control required for that task is tracking control. Since the winner of the competition carries out the relevant type of monitoring, she has tracking control, and hence her success in achieving that aim counts as a locally complete achievement. Of course, from a broader perspective, one that does not take into consideration the rules of the competition but all the relevant aspects of the shot, the winner's shot, although a complete achievement from a more local point of view, is not a complete achievement (in the broad sense of the expression 'complete achievement' defined above). As we will see, it is this broad sense that serves to define knowledge.

6.6.3 *Reply to the Problem (part II)*

Let us consider now the second part of the problem of lucky achievements. Here is the argument that Pritchard gives against the SATK:

1. STANDARD CONCEPTION OF ACHIEVEMENT: achievements are successes because of ability.
2. Artemis's success in ARTEMIS II is because of ability.
3. Therefore, Artemis's success is an achievement.
4. Artemis's success is by luck.
5. That is, some achievements because of ability are by luck.
6. SATK: knowledge is a cognitive achievement, where a cognitive achievement is a belief that is true because of cognitive ability.
7. Cognitive achievements relevantly analogous to Artemis's achievement (i.e., beliefs that are true because of cognitive ability in situations structurally similar to ARTEMIS II) are by luck.
8. Knowledge is incompatible with luck.
9. Therefore, knowledge is not a cognitive achievement.

Pritchard thinks that cases like the ones mentioned in (7) are cases structurally similar to FAKE BARNS (cases of environmental luck). In

Fake Barns, Henry's belief that there is a barn in front of him is true because of cognitive ability, but only because he has looked at the only real barn in the area. In the same way, in Artemis II, Artemis's shot is successful because of archery ability, but only because she has shot at the only target in the area that is not protected with a force shield.

My position concerning this argument is the following: I reject (1), i.e., I reject the standard conception of achievement (Standard Conception of Achievement). I rather uphold the control conception of achievement (Control Conception of Achievement). I also reject (6), i.e., I reject the standard achievement theory of knowledge (SATK), but I uphold the achievement theory of knowledge (ATK), the view that knowledge is a cognitive achievement (but not necessarily a success because of ability). I also think that (7) is false, because I do not think that there are cognitive achievements relevantly analogous to Artemis's (they are cognitive successes but not achievements). Let us see these points in more detail.

First of all, one might think that if we reconstruct the argument with the control conception of achievement, (3) and (4) would still hold, i.e., Artemis's success (in Artemis II) would still count as an achievement and it would be by luck. Suppose that we give some version of ATK in terms of control, would it not also follow that cognitive achievements relevantly analogous to Artemis's are by luck? In what follows, I will show that while Artemis's success is an achievement and by luck, lucky cognitive successes relevantly analogous to Artemis's do not count as cognitive achievements. To show that, I will give a version of ATK in terms of the control conception of achievement.

Here is the issue. Control Conception of Achievement says that achievements are successes over which the agent has either effective control, tracking control, or both types of control. ATK says that knowledge is a cognitive achievement. How can we bring together these two theses? Let us explore our options.

One possibility would be to provide a disjunctive account of knowledge of the following sort: knowledge arises just in case the agent has either (i) an epistemic analogue of tracking control (e.g., safety), (ii) an epistemic analogue of effective control or (iii) both. Is this option promising? I do not believe it is, as there does not seem to be any intuitive motivation to think that knowledge has such a disjunctive nature. There is a simpler and more plausible path that we can take.

As we have seen, Control Conception of Achievement predicts the correct ascription of achievement (given the relevant contexts) in cases where Artemis hits the mark but fails to detect the force shields and in cases where Artemis detects the force shields but fails to hit the target. For example, in the context of Artemis II, in which what matters is whether she hits the bull's eye, it is correct to say that

Artemis's shot is an achievement even though she fails to detect the force shields. On the other hand, in the context of ARTEMIS III, in which what matters is just that the participants shoot at well selected targets, it would be correct to say, regardless of whether the shots in question are accurate or not, that a shot is an achievement if the participant knows (and shows that she knows) that her target is not protected with a force shield. In neither of these two cases the relevant achievements are complete.

Now, according to the hypothesis that I introduced at the end of chapter 4 (HYPOTHESIS), nearly all (if not all) cases of true belief that is *not* knowledge are cases either of veritic luck or of veritic fortune. Since 1) knowledge is commonly thought to be incompatible with veritic luck and with veritic fortune and 2) complete achievements are, by definition, incompatible with luck and fortune, a plausible thesis to hold is that 3) knowledge is a complete cognitive achievement, i.e., not any cognitive success but a cognitive success that is indisputable from any dimension of assessment. The argument continues as follows. Since 4) a complete achievement is a success that results from a performance over which the agent has control (both of the tracking and of the effective sort), then 5) knowledge should be defined as cognitive success that results from a cognitive performance over which the agent has control, where the relevant notion of epistemic control should entail epistemic analogues of the notions of tracking and effective control.

6.7 THE CONTROL THEORY OF KNOWLEDGE

The kind of cognitive performances that we are interested in are cognitive performances whose aim is inherently the truth or, in Sosa's terms, cognitive performances in the endeavor to attain the truth. The notion of epistemic control that we seek must apply only to this kind of cognitive performances. In addition, since both tracking and effective control are required for complete achievement, we must translate in epistemic terms the notions of tracking and effective control (in chapter 5 we saw how to do it in the case of tracking control).

Why is this approach appealing, in general? Because the notion of control is broad enough to entail a variety of conditions that, when translated into epistemic conditions, accommodate the requirements and intuitions not only of the standard achievement account of knowledge (virtue epistemology in its more orthodox form), but also of modal epistemology, which is usually regarded as an alternative research program.

Thus, the control approach allows us to develop an account of knowledge that essentially holds, in keeping with the standard achievement theory of knowledge, that knowledge is a cognitive achievement (albeit a special one, viz., a complete cognitive achievement). In addition, it allows us to reinterpret the notion of cognitive achievement in a way that brings together virtue epistemology, as traditionally conceived, with modal epistemology. Moreover, modal conditions such as safety or sensitivity (safety is my preferred option) may be plausibly seen as virtue-theoretic conditions because, as we have explained in chapter 5, they capture forms of epistemic control (of the tracking sort), and epistemic control is required for cognitive achievement, and hence for knowledge, and the idea that knowledge is a cognitive achievement is the core of virtue epistemology. This holistic and conciliatory virtue-theoretic approach is what I call the *control theory of knowledge*.

6.8 THE PROBLEM OF EASY ACHIEVEMENTS

I will analyze the agential aspects of the notion of epistemic control in the next chapter. Before that, let me consider the other objection that Pritchard makes against the standard conception of achievement. According to Pritchard (*et al.* 2010: 67-8), a problem for the standard conception is that there are certain actions that are successful due to ability but do not count as achievements because of the ease with which they are brought about. An example would be the raising of one's arm:

> [I]f circumstances are really normal then there ought to be no problem with the idea that this success was *because of* the exercise of my relevant 'arm-raising' abilities. But would we naturally call the raising of one's arm in these circumstances an achievement? Intuitively, the answer is 'no'. (Pritchard *et al.* 2010: 68)

In what follows, I will argue that both the standard and the control conceptions of achievement are unaffected by this problem. But first, let us see what conception of achievement Pritchard thinks is the correct one.

6.8.1 *Pritchard's Conception of Achievement*

Here is how Pritchard conceives achievements:

- PRITCHARD'S CONCEPTION OF ACHIEVEMENT: The success of *S*'s performance is an achievement of *S* if and only if *S*'s success is because of her abilities, where the success in question involves either (i) the application of a significant level of skill or (ii) the overcoming of a significant obstacle to her success.

Examples of (i) would be the successful shot of a masterful archer in a windy environment or at great distance or Usain Bolt's victory in the Olympics 100-meters final. Examples of (ii) would be an amputee's first steps with prosthetic legs or the wiggle of a toe by a paralyzed person after struggling for hours. The raising of one's arm involves neither of these two things. Therefore, according to PRITCHARD'S CONCEPTION OF ACHIEVEMENT the raising of one's arm is not an achievement.

6.8.2 *Attainments and Achievements*

Although we would certainly use the term 'achievement' both in cases in which an obstacle is overcome and in cases of great skill, achievements that involve skill are ordinarily *not* considered equivalent to achievements that result from overcoming obstacles with great effort and no skill. The latter might be certainly considered great achievements, but only subjectively, in the sense that the overcoming of an obstacle by a person lacking the relevant skills is an achievement *for* that person. The former, by contrast, are achievements not only for the agent in question but also from a more general point of view. Since the term 'achievement' is used both in cases of great skill and in cases of great effort but no skill but we ordinarily do not treat the cases as equivalent, it would be useful to count with some terminology that would allow us to make explicit the distinction.

We can adopt John Turri's (*forthcoming*). Turri aims to avoid the problem for the standard notion of achievement that we sometimes call 'achievements' successes that result from unreliable ways of acting. On the one hand, Turri distinguishes another evaluative property of performances (in addition to accuracy, adroitness and aptness): *adequacy*. *Adequate* performances are performances whose success is because of ability *or* because of some unreliable mechanisms (or some generally unreliable way of acting). A performance that is apt (i.e., one whose success is because of ability) is therefore also adequate, but not *vice versa*. On the other hand, Turri calls *attainment* the success of an adequate performance and reserves the term 'achievement' for successes that are because of ability. Usain Bolt's victory is thus an achievement and an attainment whereas a baby or an amputee's first steps is just an attainment.

We should adopt Turri's notion of attainment for two reasons. First, the notion of attainment, so conceived, captures the sense in which a person who is merely able to φ but has no skills attains something rather than nothing when successfully φ-ing: namely to φ successfully. Second (and related to the first point), Turri's attainment/achievement distinction allows to accommodate the intuition that it is not the same to succeed with great skill (achievement and attainment) as with great effort and no skill (mere attainment). Pritchard's notion of achievement overlooks this distinction.[23] In conclusion, although

23 In ordinary discourse there is no much difference in meaning between an achievement, an attainment, an accomplishment and a fulfillment. Although the present

Pritchard's notion of achievement is not strictly wrong, it should be dropped for this reason.

6.8.3 *Two Senses of Ability*

Let us return to Pritchard's problematic example. Is it not the raising of one's arm because of one's 'arm-raising' abilities? Yes and no. Is it an achievement? No. First of all, note that the term 'ability' is ambiguous. By 'ability' we sometimes mean:

- Ability$_1$: the quality or the state of being able.

In this sense, a person who is able to raise her arm would have the ability to raise her arm. Other times, by 'ability' we mean something stronger:

- Ability$_2$: competence in doing or skill.

Skills require a dimension of proficiency relative to which one performs more or less competently. In this latter sense, a person who is able to raise her arm does not have an ability because there is nothing such as 'arm-raising' skills or competent 'arm-raisers'. We would only say that the raising of one's arm is a competent performance if acting in that way were part of an activity that is assessed along standards of proficiency (imagine a TV show in which the first participant to raise her arm may answer the one million dollar question).[24]

Is it the raising of one's arm an achievement? Only if we understand the term 'ability' in the thesis that an achievement is a success because of ability as ability$_1$. However, why should we understand the notion of achievement in that way? After all, the notion of attainment (as defined by Turri) already takes into account cases in which an agent without the relevant skills and only merely able to perform an action succeeds in performing the action. In order for one's successful φ-ing to count as an achievement, one must exhibit skill or mastery in φ-ing. Accordingly, since the raising of one's arm involves no skill, it does not count as an achievement.

In fact, contrary to what Pritchard thinks, it is the lack of ability (skill) in bringing about a success and not the easiness with which

usage of the terms 'achievement' and 'attainment' is close to the ordinary usage, it certainly diverges from it, as the achievement/attainment distinction aims to make explicit an intuitive distinction that our ordinary usage of the terms does not allow to make explicit.

24 In order to distinguish the mere quality of being able (ability$_1$) from a skill (ability$_2$) it is sometimes useful to consider whether the relevant ability could be evaluated in the context of a competition. Archery, driving and even the identification of colors are activities that can be evaluated in this way, while arm-raising, breathing, digesting or blinking cannot. There are borderline cases, though (see fn. 25). At any rate, the general idea is that in order for one to have a skill one must be able to do something that is assessable according to some pattern or ideal model of action or performance (a standard of proficiency).

it is brought about the reason we do not claim that a success is an achievement. To see this, consider for example the difference between blinking and winking (by 'blinking' I mean to close and immediately reopen both eyes and by 'winking' I mean to close and quickly reopen one eye in order to convey some meaning). Blinking at will is, in the relevant aspects, as the raising of one's arm: it is a very basic movement that one is *able* to perform just by intending to perform it and is not considered a skill.[25] Winking, by contrast, which also requires ability$_1$ (some people are unable to wink) might be considered a skill (ability$_2$), because winking might be assessed along some dimension of proficiency. When we wink, we typically intend to convey friendliness, mutual understanding, and so on, and this can be done in a more or less competent manner. Some clear cases of winking that fall under normative assessment are the following:

1. In certain card games, winking is used to pass information to one's partner. If one's wink is detected by the other team, one is obliged to tell the value of one's cards. One must wink in a proficient manner (viz., unnoticed).

2. Suppose that in the middle of a World War the future of humanity depends on a certain piece of information. Suppose that scientist *A* possesses that information, which she must pass to spy *B* at some public place. Since they do not know each other, they agree that *A* will wink two times to *B*, who will wear a red scarf. *A* is so nervous that winks three times. They fail to recognize each other. Humanity is lost.

In these cases, winking can be certainly assessed along a dimension of proficiency. In the first case, one must wink in such a way that only one's partner notices it. In the second case, winking is easier but crucial. In both cases we would say that winking successfully is an achievement. Crucially, note that the easiness with which a wink might be caused is very similar (if not the same) as the easiness with which blinks are normally caused or with which arms are normally raised. Therefore, what makes us deny that certain success like the raising of one's arm is an achievement is not the easiness with which the success is brought about, but the fact that it is not the product of skill.

6.8.4 *Competent and Basic Effective Control*

As regards the control conception of achievement, one could ask: is it not the raising of one's arm something over which one has control?

25 There are some borderline cases. To blink at will once, for example, is not a skill, as one is not competent at it, one is just able to do it. But to stop blinking for a long period of time might count as a skill, because one might be competent at it (besides being able).

Does not CONTROL CONCEPTION OF ACHIEVEMENT rule that the successful raising of one's arm is an achievement? It seems so: one has effective control over one's arm (or over its movement) just in case (i) one intends to raise one's arm, (ii) one has the disposition to raise it, (iii) one raises it and (iv) the success of the movement is because of that disposition. With a healthy body, it is undeniable that we have control over our limbs in this way.

Let us pause for a moment. Note that the second clause of EFFECTIVE CONTROL (that *A* must have the disposition to cause *B* to be in a certain state), as it stands, is silent regarding what grounds the relevant disposition. There are two possibilities:

- *A* is disposed to φ in virtue of being able to φ (i.e., in virtue of having ability$_1$ to φ).

That is, one might be disposed to φ in virtue of one's being merely able to φ without one's φ-ing being the product of skill. An example of this would be the raising of one's arm: one is disposed to raise one's arm because, unlike injured people, one is physically constituted in such a way that one is able to raise one's arm. As we have seen, however, such a power is not a skill because there is no dimension of proficiency according to which one counts as a competent 'arm-raiser'.

Yet, many basic movements are under our control in this way. So there is a sense of effective control that is not assessed by standards of proficiency (the kind of standards that we use to assess a skillful performance). For example, when one blinks at will, one has this sort of effective control over one's eyelids. People with eye twitching, eyelid tics and spasms do not have it because they are *not able* to blink or to stop blinking at will. Call this type of effective control *basic effective control*. It can be defined as follows:

- BASIC EFFECTIVE CONTROL: *A* has basic effective control over *B* if and only if (i) it is *A*'s aim that *B* is in certain state *S*, (ii) *A* is able to cause/determine *B* to be in *S*, (iii) *B* is in *S* and (iv) *B* is in *S* because of the exercise of *A*'s capability to cause/determine *B* to be in *S*.

On the other hand, *A* might be disposed to φ in this more substantial sense:

- *A* is disposed to φ in virtue of having a skill to φ (i.e., in virtue of having ability$_2$ to φ).[26]

26 But also in virtue of having a skill in which φ-ing reliably is a fundamental part of what it takes to have the skill (e.g., developing a steady hand is a fundamental part of the ability (skill) to paint figurines and miniatures). In addition, a skill may involve many sub-skills. Typically, to master a skill one has to perform reliably a variety of tasks (e.g., consider all the driving tasks required to drive). Accordingly, it is possible that a sub-skill is shared by two different skills (e.g., to have a steady hand is required for both archery and painting). See Hawley (2003) for relevant discussion on the issue.

In view of this, we may distinguish a special type of effective control that arises only when the relevant disposition to φ is a skill. This is the kind of control that safe drivers have over their cars, dexterous draftsmen over their drawings and professional soccer players over the ball. In general, this is the kind of control that we ascribe to agents engaged in practices and activities that can be assessed along some dimension of proficiency. Call it *competent effective control*:

- COMPETENT EFFECTIVE CONTROL: *A* has competent effective control over *B* if and only if (i) it is *A*'s aim that *B* is in certain state *S* (ii) *A* has a competence (a skill) that disposes her/it to cause/determine *B* to be in *S*, (iii) *B* is in *S* and (iv) *A* causes *B* to be in that state and (iv) *B* is in *S* because of the exercise of *A*'s competence to cause/determine *B* to be in *S*.[27]

As I mentioned before, our basic movements are the kind of thing over which we have basic effective control: we raise our arms, scratch our noses and move our eyes. Yet, we do not consider these things achievements. Plausibly, a success is an achievement only if it results from a skill (the success in question must exhibit some degree of proficiency). Accordingly, the type of effective control in CONTROL CONCEPTION OF ACHIEVEMENT and COMPLETE ACHIEVEMENT is competent effective control. In Pritchard's examples, there is only basic effective control.

6.8.5 *Easy Cognitive Achievements?*

There is still one problem that we must consider. Pritchard argues that the problem of easy achievements has a parallel in the case of knowledge. In the same way as he thinks that the raising of one's arm is too easy to be considered an achievement, he thinks that the true belief that the wall before one is white when one looks at a white wall in normal circumstances is too easy to be considered a cognitive achievement:

> Suppose that I form the true belief that the wall before me is white by looking at it in entirely normal circumstances. Here we have a cognitive success and the cognitive success is, intuitively, appropriately related to my relevant cognitive abilities in such a way that it is because of my cognitive ability. And yet it seems odd to think of such

27 It is an open question whether artifacts have competent effective control. All seems to depend on whether it is correct to attribute skills to artifacts. A thermostat is able to perform in a certain way in virtue of having a function, but it is clear that a thermostat does not have skills. Would we say the same thing about a very sophisticated robot? Perhaps not. For simplicity, in any case, I will ascribe competent effective control only to beings that exhibit a minimal degree of agency, such as humans or higher order animals.

> a success as an achievement on my part. (Pritchard *et al.* 2010: 69)

We have seen that the real reason we do not claim that the raising of one's arm is an achievement is not the easiness with which it is brought about but the fact that it does not involve skill. Therefore, the easiness with which the belief that the wall before one is white is formed cannot be reason to reject that it is a cognitive achievement. Are our perceptual abilities abilities in the sense of ability$_1$ (qualities of being merely able) or in the sense of ability$_2$ (skills)? I think it is quite uncontroversial to claim that they are abilities in the second sense, as they can certainly be assessed according to standards of proficiency. And I surmise that this is something that Pritchard would accept.

Now, one of the reasons that Pritchard mentions to reject that forming a so basic perceptual true belief is not an achievement is that it does not result from the exercise of significant cognitive skill. I disagree. Certain cognitive achievements are *great* cognitive achievements. For example, it is a great cognitive achievement on the part of a bird-watcher to identify at great distance a robin perched in a leafy tree. However, not all cognitive successes need such a degree of skill to be considered achievements. One may form the belief that the bird before one is a robin by watching its characteristic colors and still that cognitive success be considered an achievement on one's part. The degree of cognitive skill manifested does not change the status of one's cognitive success as an achievement, but only its significance: the former is a great cognitive achievement, the latter is a not so important one.

6.9 SUMMARY

In this chapter, I have reviewed the virtue-theoretic project in epistemology from its initial motivations to its most recent developments as a *performance-based account of knowledge*. In particular, I have analyzed the view that knowledge is a cognitive achievement (the achievement account of knowledge) and the very notion of achievement. The most common way of understanding the notion is as a success that is because of ability (the standard conception of achievement). I have proposed an alternative view: achievements are successes over which the agent has control (the control conception of achievement). As regards the achievement account of knowledge, I have analyzed its most common version (the view that a cognitive achievement and hence knowledge is a belief that is true (a cognitive success) because of cognitive ability –the standard achievement theory of knowledge–) in the light of two objections to it made by Duncan Pritchard: the problem of lucky achievements and the problem of easy achievements. I have proposed an alternative version of the achievement account of knowledge: the so-called control theory of knowledge, which conceives knowledge as a complete cognitive achievement. Complete achievements are the kind of successes that are indisputable from any dimension of assessment. In particular, I have explained, a success is a complete achievement just in case the agent has tracking and effective control over the performance that leads to that success. When a success is achieved in this way, it is neither by luck, nor by fortune. The control theory of knowledge proposes then to understand knowledge in terms of control over cognitive performance whose aim is essentially the attainment of truth. I have called this kind of control epistemic control, which is meant to be incompatible with both veritic luck and veritic fortune. The next chapters will make clear in which sense an agent has control over the formation of her beliefs in a way that makes the attainment of truth incompatible to veritic luck and veritic fortune.

For the moment, and by way of conclusion, let me note that the control theory of knowledge is not a revisionary proposal. It is rather a vindication of virtue epistemology, since it holds, in keeping with the standard achievement account of knowledge (virtue epistemology in its more orthodox form), that knowledge is a cognitive achievement (albeit a special one, viz., a complete cognitive achievement). The big advantage of the theory is that, unlike the standard account, it understands the notion of cognitive achievement in terms of control, a notion that is broad enough to accommodate the requirements and intuitions not only of virtue epistemology (chapter 7 will be precisely devoted to the connection of the notion of control with virtue-theoretic conditions for knowledge), but also, as we saw in chapter 5, of modal epistemology, which is usually regarded as an alternative research program. In this way, modal conditions such as safety or sensitivity (I opt for safety) can be seen as virtue-theoretic conditions. Nozick and Sosa's original approaches to knowledge can in this way be united under one single comprehensive framework.

7 VIRTUE EPISTEMOLOGY: EPISTEMIC CONTROL AS AGENCY

In chapter 5, we saw that the modal dimension of the notion of epistemic control can be cashed out in terms of safety (the safety condition was interpreted as an epistemic analogue of the notion of *tracking control* distinguished in chapter 3).[1] The aim of this chapter is to find an epistemic analogue of the notion of competent effective control. Recall the definition of *competent effective control* given in chapter 6:

- COMPETENT EFFECTIVE CONTROL: *A* has competent effective control over *B* if and only if (i) it is *A*'s aim that *B* is in certain state *S* (ii) *A* has a competence (a skill) that disposes her/it to cause/determine *B* to be in *S*, (iii) *B* is in *S* and (iv) *A* causes *B* to be in that state and (iv) *B* is in *S* because of the exercise of *A*'s competence to cause/determine *B* to be in *S*.

I will show that the kind of *virtue-theoretic* conditions offered by proponents of the standard achievement theory of knowledge are especially well suited to account for an epistemic analogue of the notion of competent effective control. In the final section of the chapter, I will discuss several issues related to the control theory of knowledge that will help to elucidate the proposal a bit further: the question of voluntariness and lottery cases.

7.1 SUCCESS

The first condition of COMPETENT EFFECTIVE CONTROL says that *A* has competent effective control over *B* only if it is *A*'s aim that *B* is in certain state. In the same way as Sosa thinks that the only cognitive performances that fall under the AAA structure are those that are in the endeavor to attain the truth (and not, say, to attain comfort), the notion of epistemic control needed to define knowledge is restricted only to such cognitive performances. Cognitive performances with pragmatic aims are perhaps under our control, but the kind of control we have over them (if any) cannot be called 'epistemic'. Epistemic control only arises when an agent's cognitive performance is in the endeavor to attain the truth.[2]

The third condition of COMPETENT EFFECTIVE CONTROL says that *A* does not have control over *B* unless *A* brings *B* into the state *A* aims *B* to be. Competent effective control, in particular, and effective control,

1 In this sense, epistemic control requires *local reliability*.
2 Recall that endeavors need not be explicit or conscious.

in general, requires success. If an archer intends to hit the bull's eye and fails, would we say that she has effective control over her performance? It do not think we would. We would perhaps say that the archer exerts some degree of control over her shot if, for example, she hits the target but not the bull's eye. However, that degree of control would not be sufficient to claim that she controls her shot as regards her aim of only hitting the bull's eye.

To have goals or aims is at the core of what is to be a controller and one must fulfill the goals that one attempts to accomplish with one's performances so as to have control over them. Interestingly, in the same way as we cannot say that an agent controls her performance if she fails to attain its goal, we cannot say that an epistemic agent controls her cognitive performance if she gets things wrong. In other words:

- Success: *S* controls her cognitive performance only if the resulting belief is true.

Since epistemic control requires cognitive success, our definition of knowledge (*S* knows that *p* if and only if *S* comes to believe that *p* and *S* has control over her cognitive performance) needs no separate condition on truth: there is no control without success and hence no epistemic control without truth. Any case of false belief will be automatically ruled out by the definition as a case of knowledge.

7.2 COMPETENCE

Another relevant aspect of the definition of competent effective control is that in order for *A* to control *B*, *A* must have a competence (a skill) that disposes her/it to cause or determine *B* to be in the state *A* aims *B* to be. For example, we would not say that an archer has control over her successful shot if the archer has not proved herself competent in archery. Or would we ascribe control to a novice who always fails to hit the target if she hits the bull's eye on certain occasion?

Analogously, in order for an agent to control her cognitive performance, the agent must be cognitively competent. We would not say that an agent's successful cognitive performance is under the agent's control and thus that her cognitive success (i.e., her believing truly) is an achievement of her if she lacks the relevant cognitive abilities. For example:

> Bad at Math
>
> A person attempts to do a complex calculation whose result is *n* but has no idea of how to solve that kind of mathematical problem. Suppose that this person, without having followed the correct steps, writes *n* on her notebook, just like that.

Can this person's cognitive success properly said to be an achievement of her? No, because she lacks the skills needed to solve that kind of mathematical problem. That is, she does not form her true belief that the solution is *n* in a competent way. Epistemic control, in this sense, requires epistemic competence or virtue (and hence global reliability, i.e. reliability with respect to a field of propositions and a range of circumstances):

- COMPETENCE: *S* controls her cognitive performance only if she has the relevant cognitive competences or abilities.

By including COMPETENCE in our definition of knowledge we are able to rule out as cases of knowledge cases where agents form beliefs through globally unreliable methods. Is global reliability all that is required to have epistemic competence? The definition of epistemic competence that I gave in section 6.1 says, in addition, that a cognitive disposition to form beliefs counts as an epistemic competence only if it is sufficiently integrated with other of the agent's cognitive dispositions. I will discuss the issue of cognitive integration in the next section.

7.3 COGNITIVE INTEGRATION

To begin with, certain intuitive sort of integration is required for effective control. Consider the following example:

> SUPER-ARCHER
>
> Terrible Archer is totally incompetent: she knows how to deliver arrows with a bow but she always misses the targets. Suppose that, during sleep, a neuroscientist implants a microchip in Terrible Archer's head. The microchip has the power of interacting with her muscles in the following way: whenever Terrible Archer takes aim, makes her calculations and is about to shoot, the microchip sends a signal through her nerves that subtly corrects the position of her arms and shoulders and that triggers the release of the arrow. In this way, Terrible Archer, now Super-Archer, never fails. She is very surprised that she does not. In addition, she feels like her arms are moving by themselves.

Although Super-Archer certainly has the disposition to hit targets reliably, it does not seem that *she* has control over her performance. After all, Super-Archer is astonished that she never fails and she feels like her arms are moving by themselves. A way to explain these feelings is by appealing to the fact that her new disposition is (still) not sufficiently *integrated* with the rest of dispositions that make up her agency.

In general, a disposition does not count as an ability unless it is sufficiently integrated with the rest of dispositions. Plausibly, this claim is correct even in the weak sense of ability (ability$_1$): if you reliably φ in spite of the rest of your dispositions to act, there is a sense in which you are not able to φ, at least not *adequately*. Or would we say that an agent is able to raise her arm if she has the feeling that her muscles are moving alone? Effective control (competent or just basic) requires that one's dispositions to φ are well integrated.

Importantly, the same kind of problem arises when the relevant performances are cognitive (viz., belief formation). Recall Keith Lehrer's TRUETEMP:

> TRUETEMP
>
> Suppose a person, whom we shall name Mr. Truetemp, undergoes brain surgery by an experimental surgeon who invents a small device which is both a very accurate thermometer and a computational device capable of generating thoughts. The device, call it a tempucomp, (...) is very reliable, and so [Truetemp's] thoughts are correct temperature thoughts. All told, this is a reliable belief-forming process. Now imagine, finally, that he has no idea that the tempucomp has been inserted in his brain, is only slightly puzzled about why he thinks so obsessively about the temperature, but never checks a thermometer to determine whether these thoughts about the temperature are correct. He accepts them unreflectively, another effect of the tempucomp. (Lehrer 1990: 163-164)

Here is another well-known case by Alvin Plantinga:

> BRAIN LESION
>
> There is a rare but specific sort of brain lesion (we may suppose) that is always associated with a number of cognitive processes of the relevant degree of specificity, most of which cause its victim to hold absurdly false beliefs. One of the associated processes, however, causes the victim to believe that he has a brain lesion. Suppose, then, that *S* suffers from this sort of disorder and accordingly believes that he suffers from a brain lesion. Add that he has no evidence at all for this belief: no symptoms of which he is aware, no testimony on the part of physicians or other expert witnesses, nothing. (Plantinga 1993b: 199)[3]

Finally, consider the following report given by a schizophrenic patient:

3 The agent in BRAIN LESION is supposed to form the belief that he has a brain lesion through a cognitive process that is reliable (although perhaps not *globally*).

> THOUGHT INSERTION
>
> I look out of the window and I think that the garden looks nice and the grass looks cool, but the thoughts of Eamonn Andrews [an old British TV presenter] come into my mind. There are no other thoughts there, only his... He treats my mind like a screen and flashes his thoughts onto it like you flash a picture. (Mellor 1970: 17)[4]

In these three cases, the relevant cognitive processes that lead to the relevant beliefs fail to be *cognitively integrated* and hence do not count as epistemic virtues or competences of the agents. Thus, in the same way as control in general requires that one's dispositions to φ are well integrated, epistemic control requires that one's dispositions to believe the truth are cognitively integrated. Although, strictly speaking, cognitive integration is necessary for a cognitive disposition to count as a cognitive ability, I prefer to put it (for the sake of explicitness) as a condition for epistemic control:

- COGNITIVE INTEGRATION: *S* controls her cognitive performance only if *S*'s relevant cognitive abilities are cognitively integrated.

To have a more intuitive grasp of the notion of cognitive integration, compare TRUETEMP with the following case of an agent who acquires knowledge by means of a computational device capable of generating thoughts:

> TEMPO
>
> The subject—let's call him 'Tempo'—is fitted from birth with a highly reliable device which records the ambient temperature and Tempo grows up in a culture where it is taken for granted that one consults one's temperature-recording device in order to form beliefs about the ambient temperature. (Pritchard 2010b: 146)

Like Truetemp, Tempo forms beliefs about the temperature using a temperature-recording device that has been implanted in his brain and that is globally reliable. We can even stipulate that their devices are infallible: neither Tempo nor Truetemp fail to form true beliefs about the temperature. Unlike Truetemp, Tempo's beliefs qualify as knowledge: Tempo's device, besides globally reliable or infallible, is well integrated into his cognitive system. After all, he has it implanted from birth and its use is something that is taken for granted in his society.

The intuitive difference between these two cases should suffice to have a first grasp of the notion of cognitive integration. However, we

4 Thought insertion is a psychological disorder usually associated with schizophrenia that consists in the delusion that one's thoughts (which are introspectively accessible) are being placed into one's mind by some external agent or entity.

need to state what are exactly the conditions that an agent's belief-forming disposition must satisfy in order to be integrated with other of the agent's cognitive dispositions. Desirably, the conditions offered, once generalized, will help us to explain what it takes for an agent's disposition to φ to be integrated with other of her dispositions to act (recall Super-Archer). To make sense of the idea that to control a (cognitive) performance one's (cognitive) abilities must be integrated with each other, I will build, in what follows, on work carried out by Daniel Breyer and John Greco (2008) on the notion of cognitive integration and on a further development by Tom Roberts (2012).[5]

7.3.1 *Stability and Causal Interaction*

Dispositions in general are grounded on certain constitutional features of their possessors (their constitutional bases). For example, physical properties of the atomic structure of glass ground the disposition of a bottle to break when struck. An expert archer is disposed to hit the bull's eye in virtue of having certain physical constitution: bones, muscles fibers, a visual system. The same can be said about cognitive dispositions. One is disposed to believe that there is a red surface when one looks at a red surface because one is physically and psychologically constituted in certain way. All these constitutional bases are *stable* and so are the organizational structures that make up the systems of dispositions of which these dispositions are part.

The point is that in the same way as an archer needs a stable system of dispositions to control her shots, an epistemic agent needs a stable system of cognitive dispositions to control her cognitive performances. This sort of stability is missing in Brain Lesion. In the literature, the cognitive process of Plantinga's case is usually described as a strange and *fleeting* process. The agent forms the belief that he has a brain lesion in virtue of a serendipitous cognitive process associated to his brain lesion. Thus, he is disposed to form true beliefs only as long as his brain lesion has such a side effect. It is precisely the instability of this side effect that prevents its integration with the rest of the agent's cognitive dispositions.[6] Therefore, a minimal condition for cognitive integration and hence for epistemic control is this:

- Stability: *S*'s cognitive disposition is cognitively integrated with other of *S*'s cognitive dispositions only if it is sufficiently stable.

Greco (2003) suggests another minimal intuitive condition for cognitive integration: a mark of cognitively integrated dispositions is that they *causally interact* with each other. Thus:

5 We already saw Breyer and Greco's model in section 4.1.2.
6 This is not to deny that dispositions that arise from a lesion that becomes stable may become integrated within a dispositional system. As Greco (2003: 474) points out, "a lack of health often signals a lack of integration, but this is not always the case".

- CAUSAL INTERACTION: *S*'s cognitive disposition is cognitively integrated with other of *S*'s cognitive dispositions only if it causally interacts with the other dispositions.

In BRAIN LESION, for example, there is no much reason to suppose that the cognitive process associated to the agent's brain lesion causally interacts with the rest of the agent's cognitive dispositions.[7] Thus, CAUSAL INTERACTION also explains the deficit of cognitive integration of the agent.

In any case, STABILITY and CAUSAL INTERACTION cannot be all that is required for cognitive integration. In TRUETEMP, for example, nothing prevents us from supposing that Truetemp's temperature-detector device is stable and causally interacts with the rest of Truetemp's dispositions. Yet, we do not want to claim that it is cognitively integrated.

Finally, note that stability and causal interaction requirements are also necessary for control, in general. That is, an agent has control over her performance only if her dispositions to act are stable and causally interact with each other. For example, we would not claim that a power to φ leads to controlled performances if the agent has this power only for unpredictable short intervals of time (failure of stability). We would not claim either that a patient suffering the alien hand syndrome has control over the movements of her hand (failure of causal interaction).

7.3.2 *Coherence*

Breyer and Greco (2008: 182-3) explain that another way of understanding cognitive integration is in terms of coherence. On this view, cognitive integration would be a function of coherence among beliefs.[8] More specifically:

- COHERENCE: *S*'s cognitive disposition is cognitively integrated with other of *S*'s cognitive dispositions only if the beliefs it produces cohere with the beliefs produced by the other dispositions.

---

7 Neuroscience and cognitive psychology have a rich literature on how the different sensory modalities interact to collect information about the world: what is known as *multisensory integration*. Some studies show how traumatic brain injury leads to a deficit of multisensory integration (see e.g., Sarno *et al.* 2003). See Meredith & Stein's 1993 volume *The Merging of the Senses* for an excellent introduction to multisensory integration.

8 In reality, Breyer and Greco introduce coherence as a possible way of defining the notion of cognitive ownership and cognitive ownership as a way of understanding cognitive integration. On this view, ownership would be understood as membership of a coherent web of beliefs. I prefer to put coherence simply as a condition for cognitive integration rather than as a condition for ownership and, in turn, for cognitive integration. We can in this way reserve the notion of ownership for the taking responsibility approach (which is the one that Breyer and Greco favor).

Coherence allows to rule out as cases of cognitive integration cases of extremely reliable chicken-sexers or plain spotters that believe that all their cognitive performances have been incorrect in the past.[9] The second-order beliefs of these agents conflict and hence do not cohere with their first-order beliefs about the sex of the chicks or the nationality of the plains.[10] Does Coherence rule out Thought Insertion as a case of cognitive integration as well? It seems so, as the thoughts of Eamonn Andrews that come into the patient's mind do not seem to cohere with the rest of the patient's thoughts. But what if they did? Then, we would need a stronger condition to explain why the patient's cognitive dispositions are not integrated.

We can adopt here a more substantial approach suggested by Roberts (2012: 493-4). The idea is that cognitive integration not only requires a coherent web of beliefs, but also that the whole system of stable dispositions is disposed towards coherence:

- Disposition to Coherence: *S*'s cognitive disposition is cognitively integrated with other of *S*'s cognitive dispositions only if the whole system of *S*'s cognitive dispositions is disposed to form and maintain a coherent set of beliefs.

We need to be specific on how *not* to understand Disposition to Coherence. Two points. First, the requirement that the whole system of *S*'s cognitive dispositions must be itself disposed to form and maintain a coherent set of beliefs is not a requirement on rationality. In particular, I do not mean anything like this: in order for one to be rational one must be disposed not to believe that *p* at *t* if one believes that not-*p* at *t*. What I mean, in general, is that in order for a disposition to be *cognitively integrated*, the whole system of cognitive dispositions to which the disposition belongs must tend to follow that pattern. That is to say, although there is for sure some interesting connection between cognitive integration and rationality, I am neutral on whether to be rational one's doxastic dispositions or the whole system of such dispositions must follow that general pattern. For example, I do not claim that the subject in Thought Insertion is not rational because his or her cognitive dispositions do not exhibit such a tendency. Second, I do not claim that exhibiting such a general tendency towards coherence is necessary for having beliefs. For instance, in my view the mental states of the subject in Thought Insertion count as beliefs.[11]

---

9 Cases such as Misled Chicken-Sexer, which was discussed in chapter 4.

10 In chapter 4, we saw that these are paradigmatic cases of reflective luck. The problem in section 4.3.2 was to explain in which sense this kind of cases fall under the hypothesis that all (if not all) cases of true belief that it is not knowledge are cases either of veritic luck or veritic fortune. More specifically, the problem was to explain in which sense the fact that one believes that all one's past judgments have been incorrect makes one lack epistemic control. The answer is clear now: there is a failure of cognitive integration (Coherence does not hold).

11 See Kolodny (2008) for relevant discussion on these two points.

After these preliminary remarks, consider how DISPOSITION TO COHERENCE explains some of the cases. To begin with, it seems obvious that patients suffering the thought insertion delusion do not have cognitive systems that are disposed towards coherence. DISPOSITION TO COHERENCE thus rules out THOUGHT INSERTION *as a case of cognitive integration*. In addition, as Roberts notes, this sort of condition also helps to explain why there is lack of cognitive integration in possible cases of cognitive systems whose intentional states are updated arbitrarily and happen to cohere over a particular short interval (Roberts 2012: 493-494).

§ *The concept of coherence.* We have said nothing yet about the very concept of coherence. In the specialized epistemological literature, the relation is explained in different ways (e.g., in logical or in probabilistic terms), but there is no consensus as regards the best definition of coherence.[12] For our purposes (to state a necessary condition for epistemic control), it suffices an intuitive notion of coherence that is incompatible with manifest conflict within a belief system (e.g., intuition says that, in THOUGHT INSERTION, the schizophrenic patient's beliefs and belief system are not coherent).[13]

§ *Coherence and control.* As a final point, let me briefly explain how coherence relates to control in general. In the same way as cognitive integration and thus epistemic control requires that the whole system of an agent's cognitive dispositions is disposed to form and maintain a coherent set of beliefs, control requires that the whole system of an agent's dispositions to act is disposed to perform a coherent set of actions. For example, an ambulance driver needs a balance between driving as fast as possible and not having accidents. It would be pointless for the driver to drive so fast to the point of having an accident, as it would be equally pointless to drive so safely that her slowness were cause of medical negligence. Neither reckless nor slow ambulance drivers have effective control over their performances. Another

12 See Olsson (2010) for an excellent survey on the concept of coherence (and on coherentism).

13 Nevertheless, one could demand a clearer picture of what is required for a belief system to be coherent. Here I tend to agree with Laurence BonJour (1985: 93-101) in not thinking coherence as a univocal concept (e.g., coherence as just mere logical consistency) but in thinking it as a concept with many different sides. BonJour suggests a series of conditions: (1) A system of beliefs is coherent only if it is logically consistent. (2) A system of beliefs is coherent in proportion to its degree of probabilistic consistency. (3) The coherence of a system of beliefs is increased by the presence of inferential connections between its component beliefs and increased in proportion to the number and strength of such connections. (4) The coherence of a system of beliefs is diminished to the extent to which it is divided into subsystems of beliefs which are relatively unconnected to each other by inferential connections. (5) The coherence of a system of beliefs is decreased in proportion to the presence of unexplained anomalies in the believed content of the system. Plausibly, the failure of an agent's belief system to meet any of these conditions is indicative of a deficit of cognitive integration.

example: the best basketball player in the world would show a deficit of control if during a game she scored on her own basket (the deficit in this case is perhaps not a deficit of effective control but of tracking control, of failing to monitor where her own basket is).

7.3.3 *Ownership*

Breyer and Greco's (2008) preferred way of cashing out the notion of cognitive integration is in terms of a related notion: *cognitive ownership*.[14] The general idea is that when cognitive integration fails agents do not *own* their beliefs or their cognitive processes. As we already saw in our discussion on reflective luck, this general idea may be modeled in several ways.

§ *The internalist model of cognitive ownership*. One possibility is to put it in purely internalist terms. An internalist model of cognitive ownership would (roughly) require that in order for an agent to own her beliefs (or her cognitive processes), she must have reflective access to their sources or grounds (or to the cognitive processes themselves).[15] This model, however, does not serve our purposes. Not only because a reflective access requirement is an implausible condition for knowledge (although perhaps not for justification), as we explained in chapter 4, but also because it is not a plausible condition for epistemic control. We form many of our beliefs in ways that are not transparent to us (e.g., consider the belief that a friend of yours looks like a famous actor). We nevertheless want to say that the cognitive performances that lead to such beliefs are under our control. But if reflective access is required for epistemic control, we could hardly ascribe control to belief formation. Having reflective access to the sources or grounds of one's beliefs implies that one is disposed to reflectively and hence voluntarily access the grounds of one's beliefs. The question is: how that disposition could ever be manifested in a voluntary way if we form most of our beliefs involuntarily? The internalist model of ownership is of no use to our purposes of defining knowledge in terms of epistemic control. Whatever notion of epistemic control we adopt, it must

14 Rowlands (2009) and Roberts (2012) also account for cognitive integration in terms of ownership, but they have different purposes. In particular, they attempt to support the hypothesis of extended cognition (roughly, the hypothesis that cognitive processes are not located exclusively inside the skin of cognizing organisms), a hypothesis that has gained great popularity in the last decade among philosophers of mind, cognitive psychologists and even engineers dedicated to AI. It is therefore very positive that we include in our theory of knowledge a notion of cognitive integration that is not only compatible with this hypothesis but that also serves to make sense of it. See Pritchard (2010b) for an interesting exploration of how the hypothesis of the extended cognition might be accommodated by our current theories of knowledge.

15 See Bernecker 2008 for further discussion on this model.

not entail (direct) voluntariness in belief formation (I will address the question of voluntariness in section 7.6.1).[16]

§ *Ownership and taking responsibility.* The challenge then is to find conditions for cognitive ownership that are neither too demanding (as a reflective access or awareness requirement is) nor too deflated (as Stability and Causal Interaction are). That is, we neither want that cognitive ownership, which is necessary for cognitive integration and, in turn, for epistemic control, makes the latter notion entail inadmissible forms of doxastic voluntarism, nor we want that it regards as cognitively integrated cognitive processes that look like sub-personal mechanisms rather than agent abilities. To give an example of the latter point: as Lehrer describes it, Truetemp's belief-forming process seems to operate more like, say, digestion than as a source of knowledge that the agent recognizes as such (both Truetemp's device and the dispositions responsible for digestion meet stability and causal interaction requirements).

Breyer and Greco's preferred model of cognitive ownership satisfies these *desiderata*. The general idea of this model is that one owns one's beliefs when one *takes responsibility* for one's mechanisms of belief formation. They borrow the idea (and the model in general) from John Martin Fischer and Mark Ravizza's (1998) account of moral responsibility. Since they extrapolate important aspects of Fischer and Ravizza's view to the epistemic case, it is worth pausing for a moment to present Fischer and Ravizza's account.

16 One may think that we should reject for the same reasons the reflective endorsement model of ownership that Breyer and Greco (2008: 179-180) sketch as an alternative to their preferred model. According to this alternative model, in order for a belief (or a cognitive disposition) to be owned by an agent, the agent must have a *perspective* on it. Breyer and Greco claim that this is closely related to Sosa's virtue perspectivist approach to reflective knowledge and subjective justification (Sosa 1991). For Sosa, to have a perspective means that the agent "must have some awareness of one's belief and its source, and of the virtue of that source in general and its specific instance" (Sosa 1991: 292). One may think that ownership defined in terms of a so understood concept of perspective is just the internalist notion of ownership. However, Sosa's awareness condition is more liberal than the sort of awareness conditions that the internalist typically puts forward. One the one hand, unlike the internalist, Sosa thinks that awareness of the source does not require great precision and detail. For example, some grasp is required, "even if it remains sketchy and generic" (Sosa 1994: 30), of the circumstances under which the source is reliable and the sort of propositions with respect to which it typically yields true beliefs (this point has been challenged by Greco (2004) nevertheless). On the other hand, Sosa claims that one's perspective needs not be conscious: "*Conscious* reflection on the spot is not required, however, since a second-order perspective can work beneath the surface of consciousness" (Sosa 2004: 291-292). So interpreted, the reflective endorsement model of belief ownership shares, as we will see, important features with Breyer and Greco's 'taking responsibility' model (although Breyer and Greco explicitly reject that this model requires anything like a perspective of the agent). In any case, it is Breyer and Greco's model that I favor most.

§ Fischer and Ravizza on control and ownership. To begin with, Fischer and Ravizza's theory of moral responsibility is based on the general idea that responsibility is associated with control, an idea that fits in with the present approach to knowledge. The kind of control that moral responsibility requires is what they call *guidance control*. They understand guidance control in terms of two notions: the agent's *ownership* of the mechanism that issues in the relevant behavior and the reasons-responsiveness of that mechanism. Here, I will only pay attention to the former idea. They think that an agent owns certain mechanism when she *takes responsibility* for the behavior that arises from that kind of mechanism. The most typical way in which this takes place is, according to Fischer and Ravizza, one in which the agent does not explicitly *reflect* on the relationship between causal determinism and moral responsibility. To illustrate the point, they tell the following story about moral education:

> As a child grows up, he is subject to moral education (imperfect as it may be). The child's parents—and others—react to the child in ways designed (in part) to get the child to take certain attitudes toward himself: to view himself in certain ways. Partly as a result of moral education, the child typically acquires the view of himself as an *agent*, in at least a minimal sense. That is, he sees that upshots in the world depend on his choices and bodily movements. Further, the child comes to believe that he is a fair target of certain responses—the "reactive attitudes" and certain practices, such as punishment—as a result of the way in which he exercises his agency. We claim that it is in virtue of acquiring these views of himself (as a result of his moral education) that the child *takes responsibility*. More specifically, it is in virtue of acquiring these views that the child takes responsibility for *certain kinds of mechanisms*: practical reasoning, non-reflective habits, and so forth. (Fischer & Ravizza 2000: 442)

There are two important points in this story. On Fischer and Ravizza's view, taking responsibility "is a matter of having certain (dispositional) beliefs about oneself (and having acquired those beliefs in appropriate ways)" (Fischer and Ravizza 2000: 443). If this is correct, then the process of taking responsibility needs *not* be *explicit*, *conscious* or *reflective*. This is important because we do not either want that, in the epistemic case, the process of taking responsibility for belief is necessarily conscious or reflective (although it might be).

The second point in Fischer and Ravizza's story is that the process of taking responsibility is *genuinely historical*. As they explain, in order for an agent to be morally responsible for her behavior, the process of taking responsibility must take place at some point in the past, i.e., before the behavior is produced by the relevant reasons-responsive mechanism. Notice that all the properties of the notion of cognitive integration presented so far (stability, causal interaction, coherence) are ahistorical. It is therefore important that we include a historical element in our account of cognitive integration, because, as we will see, only by taking into consideration the agent's historical development we will be able to explain certain cases of failure of cognitive integration.

By way of illustration, Fischer and Ravizza present a very useful analogy. A perfect fake of a Picasso is not a Picasso even though the expert's eye can spot no difference: being a Picasso is a genuinely historical phenomenon. In the same way, they point out, two agents may be disposed to act in the same manner and may have very similar (perhaps identical) physical and psychological attributes and yet one agent be morally responsible for her behavior and the other not. The reason: with the passing of time, one takes responsibility for the mechanism that produces her behavior while the other does not.

To illustrate this reason, Fischer and Ravizza present some cases of agents whose brains have been manipulated in ways that are not transparent to them and that compel them to behave in certain ways. Curiously, we have considered cases (Truetemp and Tempo) of agents (Truetemp and Tempo) who have been implanted with temperature-recording devices capable of generating thoughts. Truetemp has no idea of why he is so obsessed with the temperature and, as we will see, he does not *own* his thoughts about the temperature, in the sense that has not *yet* taken responsibility for the way he forms them. Unlike Truetemp, Tempo took responsibility for his thoughts long time ago: he was born with the device and he lives in a society where the use of this device is acknowledged and accepted.[17]

More needs to be said on what is exactly required for taking epistemic responsibility for belief formation. But as a general point, it is now sufficiently clear that we need to adopt some condition on cognitive integration that takes into account historical elements of the agent's development.

§ Breyer and Greco's model of cognitive ownership. Inspired by Fischer and Ravizza's account of moral responsibility, Breyer and Greco (2008: 181) suggest the following three conditions for taking responsibility for belief and thus for cognitive ownership: (i) *S* must recognize herself as the source of her beliefs (such as perceptual beliefs); (ii) *S* must accept that she is fairly credited with having certain beliefs; (iii) *S* must base her beliefs about herself on appropriate grounds.

They do not give much details on how to understand these conditions or on how they are to be applied to the cases. Roberts (2012) fills out the details. The first point to note is that, in the same way as in Fischer and Ravizza's view the process of taking responsibility

17 The same story applies to BonJour's clairvoyants.

for one's behavior needs *not* be *explicit*, *conscious* or *reflective*, taking responsibility for a belief does not require that one explicitly, consciously or reflectively thinks about how one has acquired it, i.e., no Cartesian self-examination is needed. How is then condition (i) to be understood? Roberts gives the following explanation:

> [R]ecognizing that one possesses a particular mode of belief-acquisition—for example, a perceptual channel—can be understood as a matter of coming to grasp that one is sensitive to certain perceptually-accessible properties under certain circumstances; that one can learn about features of the world and make use of this knowledge in the service of thought and activity. (Roberts 2012: 497)[18]

Roberts thinks that recognizing oneself and one's belief-forming mechanisms as a source of beliefs in this way is the basis upon which one accepts that one is fairly credited with having certain beliefs (condition [ii]).[19] What does it mean then that one must base one's beliefs about oneself on appropriate grounds (condition [iii])? Here is what Roberts says in this regard:

> One gains beliefs about one's own cognitive powers, we might say, by exercising one's abilities in the world; by testing and co-ordinating one's mechanisms of detection, storage, and behavior, and coming to an unreflective, practical understanding of one's capabilities. In the case of our innate perceptual capacities, this kind of familiarity will be gained over the normal course of development, mediated by exploration of, and feedback from, the physical environment. But an individual could also come to achieve responsibility for a novel method of belief-acquisition, and its products, over time. (Roberts 2012: 497)

Recognizing oneself as the source of one's beliefs, accepting that one is fairly credited with having certain type of beliefs and basing one's beliefs about oneself on appropriate grounds is a process that requires

18 In footnote 16, I argued that Sosa's reflective endorsement model of ownership is not much different to Breyer and Greco's. There is a point of convergence here. As I explained, Sosa does not think that the relevant perspective needs to be conscious. He thinks that to have a perspective one needs some sketchy and generic grasp of the circumstances under which a particular mode of belief-acquisition is reliable and of the sort of propositions with respect to which it typically yields true beliefs. This seems very similar to the claim that to recognize oneself as the source of one's beliefs one must grasp that one is sensitive to certain perceptually-accessible properties under certain circumstances and that one can learn about features of the world and make use of this knowledge in the service of thought and activity.

19 It is not easy to specify what is meant by recognition in this context. It might mean implicit belief that one's belief-forming mechanisms are a source of beliefs. Perhaps, it minimally means a disposition to believe that explicitly. In any case, by recognition in this context it is not meant knowledge.

a gradual and historical development. In most cases, as Roberts notes, we take responsibility over our cognitive powers by making them interact with the physical environment and by learning from that interaction. For other types of mechanisms of belief formation (e.g., epistemic devices such as microscopes), we may need someone to guide us in order to recognize ourselves as competent epistemic agents in the use of those epistemic devices.

In conclusion, only once we have taken responsibility for our belief-forming mechanisms and powers (either by gradual interaction with the environment or with the help of someone else), it may be properly said that they are cognitively integrated. Thus:

- OWNERSHIP: *S*'s cognitive disposition is cognitively integrated with other of *S*'s cognitive dispositions only if:
  - (i) *S* recognizes herself as the source of the beliefs produced by that disposition,
  - (ii) *S* accepts that she is fairly credited with having certain beliefs (the kind of beliefs that the disposition produces),
  - (iii) *S* bases her beliefs about herself on appropriate grounds (e.g., as a result of an appropriate learning process).

§ *Cases*. Let us consider again TRUETEMP and TEMPO. One of the reasons why TRUETEMP must be ruled out as a case of cognitive integration is that it does not satisfy the historical element needed to satisfy the conditions of OWNERSHIP. As Lehrer describes the case, Truetemp seems to lack the sort of unreflective, practical understanding of his detection ability, the kind of understanding that one can only achieve by gradually interacting with the environment under different conditions or by being involved in a learning context. In addition, and as a consequence of that lack of understanding, Truetemp is puzzled about his obsession about the temperature. This kind of feelings indicate that he does not properly recognize himself as the source of his beliefs. As Roberts explains:

> Our unreflective recognition of ourselves as the source of beliefs is evidenced by our lack of *surprise* at coming to gain new information through familiar channels, our dispositions to deploy our belief-forming faculties when it is appropriate to do so, and our failure to subject their deliverances to lengthy reflective scrutiny. (Roberts 2012: 497)

By contrast, TEMPO is a clear case of cognitive integration. Since Tempo is fitted with the device from birth and he has grown up in a culture where its use is a socially accepted practice, we may suppose that Tempo has engaged in the kind of social practices that make one master the device and thus recognize it as one more perceptual channel.

Consider again the cases of agents with unusual cognitive abilities discussed in chapter 4 (chicken-sexers, plain spotters, temperature detectors, clairvoyants). We can say now, in the light of OWNERSHIP, that the beliefs they form through reliable unusual cognitive abilities might qualify as knowledge only if they reflectively or unreflectively recognize themselves as the source of those beliefs, which they must endorse without conflict, and only if this recognition *has arisen* (note the past tense) from regular interaction with the world under a range of different conditions or from some guided learning process.

There is one last kind of case that we need to consider: testimony. One may argue that in testimony cases one cannot properly recognize oneself as the *source* of one's beliefs, because the origin of one's beliefs is another agent. The reply that Roberts gives to this objection is to restrict what counts as 'source' to the *latter stages* of belief formation. In his words, sources are "the proximal sensory and reasoning mechanisms that lead to the stable possession of a representational state" (Roberts 2012: 498); in the testimony case, what counts as source are "the skillful deployment of the auditory system, the mechanisms of attention and the psychological faculties of evaluation" (*ibid.*).

§ Ownership and control. We can conclude that OWNERSHIP is a plausible condition for cognitive integration and, in turn, for epistemic control. Is something like OWNERSHIP a plausible condition for control in general as well? I think it is. Consider again the case with which we began this section: SUPER-ARCHER. Super-Archer never misses a target. However, we cannot say that her performances are under her control. Although competent (she always hits the bull's eye), she succeeds because of an implanted microchip that is not integrated with the rest of her powers. This is evident because Super-Archer is very surprised that she does not fail and, in addition, she feels like her arms are moving by themselves. In this way, she does not recognize herself (yet) as the source of her successful shots. She has not taken responsibility for her new power to shoot.

7.3.4 *Cognitive Integration Defined*

The thesis held in this section is that epistemic control requires cognitive integration [COGNITIVE INTEGRATION]. We have seen a variety of necessary conditions for cognitive integration. Plausibly, they are also jointly sufficient. *S*'s cognitive disposition is cognitively integrated with other of *S*'s cognitive dispositions if and only if:

- [STABILITY] it is sufficiently stable,
- [CAUSAL INTERACTION] it causally interacts with the other dispositions,

- [COHERENCE] the beliefs it produces cohere with the beliefs produced by the other dispositions,
- [DISPOSITION TO COHERENCE] the whole system of *S*'s cognitive dispositions is disposed to form and maintain a coherent set of beliefs,
- [OWNERSHIP (i)] *S* recognizes herself as the source of her beliefs,
- [OWNERSHIP (ii)] *S* accepts that she is fairly credited with having certain beliefs (the kind of beliefs that the disposition produces), and
- [OWNERSHIP (iii)] *S* bases her beliefs about herself on appropriate grounds (e.g., as a result of an appropriate learning process).

I exclude from the definition an (internalist) reflective access requirement (because of its incompatibility with epistemic control)[20] as well as an (externalist) reflective endorsement requirement (because of its similarity with Breyer and Greco's model of taking responsibility).[21]

7.4 SUCCESS *because of* COMPETENCE

In this section, we will see how condition (iv) of COMPETENT EFFECTIVE CONTROL (*S* has control over her performance, only if its success is because of the agent's abilities) can be plausibly seen as a condition for epistemic control as well. The reason to add a condition like (iv) to the definition of effective control (not just to the definition of competent effective control) is that in some cases an agent possesses the abilities needed to φ but her successful φ-ing is due to some factor that has nothing to do with her, which makes her lose control over her performance. For example, in order for an archer to have effective control over her shot the success of the shot must be due to or because of the exercise of her archery abilities and not because of the wind. In the same way:

- SUCCESS BECAUSE OF COMPETENCE: *S* has control over her cognitive performance only if *S*'s cognitive success is *because of* the exercise of *S*'s cognitive competences or abilities.

The standard achievement account of knowledge (SATK) (virtue epistemology in its most orthodox form) defines knowledge precisely in these terms:

- *S* knows that *p* if only if *S*'s cognitive success (i.e., *S*'s believing truly) is (sufficiently) due to/because of cognitive competence or ability (i.e., epistemic virtue).[22]

---

20 See section 7.3.3.
21 See footnotes 16 and 18.
22 Greco (2010), Sosa (2007/2009; 2011) and Zagzebski (1996) defend this general view.

The real dispute among virtue epistemologists (not just among proponents of the achievement account of knowledge) is on how to understand the technical 'because of'. Although some regard it as primitive (e.g., Zagzebski 1996), virtue epistemologists are basically divided into two camps:[23] those who understand the relation in terms of *explanatory salience* or *creditability* (e.g., Greco 2010; Pritchard *et al.* 2010, 2012b, 2012c; Riggs 2002, 2009; Sosa 2007) and those who interpret it in terms of *manifestation of competence* (e.g., Kelp 2012; Turri 2011, *forthcoming*; Sosa 2010, 2011).[24] How should we read the 'because of' in Success Because Of Competence? Let us analyze the pros and cons of each proposal.

7.4.1 *Explanatory Salience / Creditability*

The general idea concerning this reading of the 'because of' relation is this:

- Creditability: If *S* knows that *p*, *S*'s cognitive success (i.e., *S*'s believing truly that *p*) is *explained* (with some degree of *salience*) by the exercise *S*'s cognitive abilities.

Alternatively, authors who endorse this reading think that this is an alternative way to put it:

- Creditability: If *S* knows that *p*, *S*'s cognitive success is (to some degree) *creditable* (or *attributable*) to *S*'s cognitive abilities.

Two points of clarification are in order.

1. *The notion of cognitive success*. The notion of cognitive success is twofold: it means *believing truly*. Accordingly, the exercise of cognitive ability might explain two things: the existence of the belief (why the agent *forms* the belief that *p*) or its correctness (why the agent comes to believe the *truth* concerning *p*).[25] When a belief is knowledge, the exercise of the agent's cognitive abilities explains both things (or both things are creditable to their exercise). Therefore, there might be cases in which there is no knowledge because the exercise of the agent's cognitive abilities explains the existence of the relevant belief without explaining its correctness (i.e., its 'hitting' on the truth). Note that:

> Something may explain the existence of a certain entity, however, without even partially explaining why it has a given property. (Sosa 2007: 65)

23 Here I follow Kelp (2012).
24 Sosa seems to have changed his view: in his 2007 book he explicitly reads the 'because of' relation in explanatory/creditability terms; in his more recent book (2011), he opts for the manifestation reading.
25 Sosa (2007: 33, 95) makes this distinction.

Sosa gives a nice example: the existence of a Volvo might be explained by its being made in a Volvo factory but that does not necessarily explain why it is defective now. Sosa's diagnosis of Gettier-style cases is analogous: only the existence of the beliefs of Gettier subjects is creditable to their epistemic competences; their correctness is explained by accidental factors, which is the reason why Gettiered beliefs do not qualify as knowledge.

2. *Creditworthiness vs. creditability*. Some authors (e.g., Greco 2003), Riggs 2002, 2009 and plausibly Sosa 2007) endorse the so-called credit view of knowledge, according to which if *S* knows that *p*, then *S deserves credit* for truly believing that *p*. This principle is often considered equivalent (not only by them but also by some of their critics [e.g., Lackey 2007, 2009]) to the principle that if *S* knows that *p*, *S*'s cognitive success is *creditable* to *S*'s cognitive abilities. However, thinking that the two principles are equivalent is a mistake. As Pritchard (2012b: 264, fn. 26) correctly points out, there is a difference between one's cognitive success being creditable to one's cognitive abilities and one's cognitive success being of credit. Let me give an example: it is certainly creditable to *S*'s memory skills having memorized all the names that appear on the phone book, but the unusefulness of such a cognitive achievement does not make it creditworthy, for example when compared to the discovery of some cure for some mortal disease.[26]

CREDITABILITY leads to two views on knowledge, the difference between which lies in the degree of explanatory salience or creditability required for knowledge. Let us see them in more detail.

7.4.1.1 *Full Creditability*

The strongest view, which is endorsed for example by John Greco (2010), says that:

- FULL CREDITABILITY: *S* knows that *p* if and only if *S*'s cognitive success is *primarily creditable* to *S*'s cognitive abilities.

Alternatively:

- FULL CREDITABILITY*:* *S* knows that *p* if and only if the exercise of *S*'s cognitive abilities is the *most salient* factor in the total set of factors that explain *S*'s cognitive success.

§ *Pros*. FULL CREDITABILITY is especially suitable to eliminate intervening veritic luck. It also rules out cases of veritic fortune as cases of

26 History provides more examples: does a scientist deserve credit for direct intellectual contribution to the construction of the atomic bomb? Does a physician deserve credit for discovering the limits of human body by experimenting on human subjects? Yet, the knowledge gained through their research is creditable to them.

knowledge, but only those that are modeled on cases of intervening luck.

1. *Intervening veritic luck*. FULL CREDITABILITY explains why there is not knowledge in cases of intervening luck. For example, in COINS (Gettier's famous case of the job applicant who comes to believe truly but luckily that the man who will get the job has ten coins in his pocket)[27] the most salient factor that explains why that person gets things right is that he is (unbeknownst to him) the person who finally gets the job and that he has, luckily, ten coins in his pocket. While the exercise of his cognitive abilities is the most salient factor in the explanation of why his belief *exists*, it is not a salient factor in the explanation of why he comes to form a *correct* belief. Alternatively, it is not creditable to his cognitive abilities that he gets things right.

2. *Veritic fortune*. FULL CREDITABILITY also explains why there is not knowledge in several cases of veritic fortune. By way of illustration, consider TEMP (the case of the agent who forms beliefs about the temperature by consulting a thermometer that fluctuates randomly but whose beliefs are always true because a hidden agent diligently changes the temperature of the room so as to match the readings of the unreliable thermometer).[28] The diligent intervention of the hidden agent is clearly what explains why the agent (Temp) gets things right and, as per FULL CREDITABILITY, the reason his beliefs do not qualify as knowledge. Nevertheless, FULL CREDITABILITY incorrectly judges that there is knowledge in cases of veritic fortune constructed on the basis of cases of environmental luck such as FAKE BARNS, precisely because the latter are problematic for the view.[29]

§ *Cons*. In the literature, there are three type of troublesome cases for FULL CREDITABILITY: testimony cases, cases of environmental luck and cases of extended cognition. Jennifer Lackey (2007, 2009) is particularly concerned with the difficulty that this view has in accounting for testimonial knowledge. Duncan Pritchard (*et al*. 2010; 2012b) is more concerned with cases of environmental luck but, as we will see in a moment, he also presents a devastating dilemma against views based on FULL CREDITABILITY using both types of cases. Finally, Krist Vaesen (2011) argues that there is a tension between FULL CREDITABILITY and extended cognition.

1. *Testimonial knowledge*. Consider the following case of testimonial knowledge:

27 See section 4.3.1. I say 'luckily', but as I explain in that section the case might be also considered a case of veritic fortune depending on how it is described.
28 See section 4.3.
29 We will see a case of veritic fortune modeled on a case of environmental luck in chapter 8. See section 8.1.

Chicago Visitor

> Having just arrived at the train station in Chicago, Morris wishes to obtain directions to the Sears Tower. He looks around, approaches the first adult passerby that he sees, and asks how to get to his desired destination. The passerby, who happens to be a lifelong resident of Chicago and knows the city extraordinarily well, provides Morris with impeccable directions to the Sears Tower by telling him that it is located two blocks east of the train station. Morris unhesitatingly forms the corresponding true belief. (Lackey 2009: 29)

The problem for Full Creditability is that in cases of testimonial knowledge the cognitive successes in question are not *primarily* creditable to or *fully* explained by the relevant cognitive abilities. In Chicago Visitor, the correctness of Morris's belief that the Sears Tower is located two blocks east is not primarily creditable to Morris's cognitive abilities, because part of the correctness is creditable to the passerby's. In other terms, the most salient factor in the total set of factors that explain why Morris believes the truth is not the exercise of his cognitive abilities, since the passerby's cognitive abilities are a salient factor (if not the most salient) of that explanation. Therefore, Full Creditability is too strong, as it rules out cases of testimonial knowledge as cases of knowledge.

2. *Environmental veritic luck*. By contrast, cases of veritic environmental luck seem to prove that Full Creditability is too weak. In Fake Barns, for example, Henry forms the true belief that the object in front of him is a barn by looking at a genuine barn. His visual faculties are very reliable; the barn is few meters from him; the light conditions are excellent; he looks at it and forms the belief in question. Nothing prevents us from thinking that the correctness of Henry's belief is primarily creditable to the exercise of his visual faculties. That is, it is reasonable to think that the most salient factor in the total set of factors that explain why Henry comes to believe the truth is the fact that he exercises visual ability. Yet, most epistemologists think that Henry does not know that the object in front of him is a barn because he could very easily have looked at a barn facsimile. Therefore, Full Creditability seems too weak, as it judges that there is knowledge in some cases of knowledge-undermining luck.

There might be ways of accommodating testimonial knowledge and of ruling out environmental luck.[30] However, the big problem for Full Creditability is not to provide independent solutions to each problem separately but to solve both problems while still offering a unified account of knowledge. This point has been forcefully

30 See Greco (2010) and Riggs (2009).

made by Pritchard (e.g., 2012b), who suggests a very simple but devastating dilemma: in order to rule out cases of environmental luck as cases of knowledge the defender of Full Creditability is pushed to strengthen her view, while to account for testimonial knowledge she is compelled to weaken it. Consequently, the cases pull the theory in opposite directions, at the obvious risk of splitting it into two dissociated parts.

3. *Extended cognition.* Vaesen (2011) presents an interesting case of human attention increased by technology. He describes an airport baggage scanner (System 2) which unlike its predecessor (System 1) inserts digital fictional threat images of guns, knives or bombs in order to make the operator believe that these items are actually packed inside the passenger's bag. Whenever the operator notices any of these items, a false alarm message pops up. System 2 produces a great increase in vigilance level. Imagine now the following situation:

> Sissicase
>
> Sissi has been a baggage inspector all her life. She used to work with an old-fashioned System 1, but since 9/11, the airport she is working for introduced a System 2. (...) Currently Sissi is inspecting a piece of luggage which contains a bomb. She notices and forms a true belief regarding the contents of the suitcase. As such, the bomb is intercepted and a catastrophe prevented from happening. (Vaesen 2011: 523)

Against Full Creditability, Vaesen argues that the most salient factor in the explanation of why Sissi comes to believe truly (and know) that there is a bomb is not the exercise of her detection abilities, because the fact that she uses System 2 is also a salient factor in that explanation. Therefore, he thinks, Full Creditability is too strong to account for cases of knowledge in which the relevant belief-forming processes are cognitively extended.

A possible reply by Greco (or any defender of Full Creditability) could be the following. The use of System 2 would explain Sissi's cognitive success in the same way as the use of a caffeine pill would: both enhance Sissi's attention. However, it seems that, in doing so, they explain Sissi's cognitive success only indirectly. In particular, System 2 and the caffeine pill would explain why Sissi's is in a position to believe the true proposition that she actually believes and, therefore, why she is in a position to know. But explaining why one is in a position to know a proposition (or why one is capable of knowing it) is one thing; another thing, is to explain why one comes to form a true (rather than a false) belief about that proposition, i.e., why one's belief ends up 'hitting the truth'. In this way, Sissi's cognitive success would be fully creditable to her cognitive abilities and the fact that she is in a good epistemic position would be creditable to SYSTEM 2. Perhaps Greco (or the defender of FULL CREDITABILITY) could adopt this kind of reply to cases of knowledge by extended cognitive processes. However, he would still have to explain the tension existing between cases of testimonial knowledge and cases of environmental luck. In fact, this tension is one of the reasons that have made Greco change his view (Greco 2012).[31]

7.4.1.2 *Partial Creditability*

As we have seen, Pritchard (*et al*. 2010, 2012b) thinks that FULL CREDITABILITY is too strong to account for testimonial knowledge, but he thinks that a successful theory of knowledge should include some virtue-theoretic condition in terms of creditability. He endorses the following virtue-theoretic condition:[32]

- PARTIAL CREDITABILITY: If *S* knows that *p*, *S*'s cognitive success is *to a significant degree* creditable to *S*'s cognitive abilities.

Alternatively:

- PARTIAL CREDITABILITY: If *S* knows that *p*, the exercise of *S*'s cognitive abilities is a *salient* factor in the total set of factors that explain *S*'s cognitive success.

§ *Pros*. There are three types of cases that PARTIAL CREDITABILITY accommodates correctly:

31 His reading of the because of relation is now a pragmatic one: «A success is attributable to S's ability [i.e., it is *because of* S's ability] just in case S's ability contributes to that success in the right way, where ''in the right way'' means ''in a way that would regularly serve relevant purposes''» (Greco 2012: 14). Greco defines knowledge accordingly: *S* knows that *p* just in case (1) *S*'s believing that *p* is produced by an exercise of intellectual ability of the right sort (i.e., "of a sort that would regularly serve relevant informational needs, both local (actual) and global") and (2) *S*'s belief being so produced contributes (in the right way) to *S*'s having a true belief, where the expression 'in the right way' means "in a way that would regularly serve relevant informational needs, both local and global" (Greco 2012: 19). The view seems to be open to objections of the following sort: *S* has a cognitive ability *A* of the right sort (i.e., one that regularly serves relevant informational needs), yet the method by which *S* forms her belief, of which *A* is part, is very unusual and, intuitively, does not *regularly* serve relevant informational needs (neither locally, nor globally); yet, *S* knows that *p*. For example, imagine that you have excellent eyesight. Imagine also that you get access to your visual evidence by a random method that delivers misleading evidence half of the times and correct evidence the other half of the times. Suppose that you come to know a proposition when the evidence is not misleading. (1) is satisfied, since you believe the truth by an exercise of visual ability "of a sort that would regularly serve relevant informational needs, both local (actual) and global". However, (2) does not seem to be satisfied in this case, because the way the belief is produced does not contribute "in a way that would *regularly* serve relevant informational needs, both local and global" to your coming to believe the truth.

32 As we will see in section 8.2, Pritchard also accepts safety as necessary for knowledge.

1. *Testimonial knowledge*. Consider CHICAGO VISITOR. Pritchard thinks that, although it is too strong to claim that Morris's cognitive success is *fully* creditable to Morris's cognitive abilities, it is plausible to claim that it is *partially* creditable to them. After all, he argues, we may suppose that Morris would not have asked anyone, but someone looking like a reliable informant (e.g., an adult rather than a child). Therefore, Morris displays significant competence in picking a reliable informant and although the passerby's cognitive abilities are of course a salient factor in the explanation of why Morris gets things right, Morris's cognitive abilities are also a salient factor in such an explanation (but not the most important one).

2. *Extended cognition*. The same reasoning can be made about cases of extended cognition. PARTIAL CREDITABILITY does not rule out SISSI-CASE as a case of knowledge. After all, although SYSTEM 2 increases Sissi's attention, Sissi still needs to exercise her detection abilities and consequently her cognitive success is partially creditable to them.

§ Cons. We saw that FULL CREDITABILITY cannot exclude environmental luck. In the same way:

1. *Environmental veritic luck*. PARTIAL CREDITABILITY is not able to rule out cases of environmental luck (e.g., FAKE BARNS) as cases of knowledge. Nevertheless, this is something that Pritchard is happy to accept, and the reason why he does not define knowledge solely in terms of creditability (recall that PARTIAL CREDITABILITY only states a necessary condition).

2. *Intervening veritic luck.* In addition, although PARTIAL CREDITABILITY might rule out some cases of veritic luck as cases of knowledge (especially most Gettier-style cases, as in most of them the fact that the agent comes to believe the truth is solely explained by some accidental factor not by the agent's cognitive abilities), there are some cases of veritic luck (some non-standard Gettier-style cases) in which the consequent of PARTIAL CREDITABILITY holds. We will see an example in chapter 8 (section 8.2.2).

3. *Testimonial knowledge*. According to Pritchard, CHICAGO VISITOR shows that in cases of testimonial knowledge the correctness of an agent's belief is always partially creditable to her cognitive abilities. Lackey (2009) casts doubt on this claim. She gives the following example:

> I come to believe that the bird on the tree is a Harris's hawk, in part because I am able to identify it as a hawk —rather than, say, an eagle or a falcon— and in part because your expert ornithological testimony enables me to specifically classify it as a Harris's hawk. (Lackey 2009: 40)

And she argues:

> There is a perfectly reasonable sense in which the truth of my Harris's hawk belief here seems to involve *partial correctness* that is *fully attributable* to me. The generic hawk part of my belief is fully attributable to me, and the Harris's part of my belief is fully attributable to your expert testimony. (*ibid*.; emphasis mine)

Given this alternative explanation, why should the fact that this testimony-based belief is knowledge be explained in terms of full correctness that is partially creditable to cognitive ability rather than in terms of partial correctness that is fully creditable to cognitive ability? I agree with Lackey that to choose one rather than the other option strikes as an *ad hoc* move. As we will see next, the manifestation reading of the 'because of' relation can avoid this kind of problems concerning degree constraints on creditability or explanatory salience.

7.4.2 *Manifestation (Aptness)*

While CREDITABILITY considers the 'because of' relation an *explanatory* relation, the following interpretation thinks of it as a *metaphysical* relation. On this reading, to say that if *S* knows that *p*, *S*'s cognitive success is *because of* the exercise of her cognitive abilities means that:

- APTNESS: If *S* knows that *p*, *S*'s cognitive success (i.e., *S*'s believing truly that *p*) *manifests* epistemic competence (or cognitive ability).[33]

33 A note on the terminology adopted. Turri (2011) explicitly introduces this reading as an alternative to the creditability reading and, in particular, as an alternative to Greco's condition for knowledge that *S* knows that *p*, only if *S*'s cognitive success is primarily creditable to cognitive ability. But he also introduces it as an improvement of Sosa's aptness condition and, in fact, he gives a different name to APTNESS: 'adeptness'. Turri's introduction of new terminology is understandable, as Sosa's notion of aptness is anything but univocal. Sometimes, Sosa speaks of aptness as cognitive success *through* the exercise of epistemic competence (reading one). But on other occasions he conceives aptness as cognitive success *attributable* (i.e., *creditable*) to epistemic competence (reading two). Yet other times he takes it to be cognitive success that *manifests* epistemic competence (reading three). The first reading appears in Sosa's first papers on virtue epistemology (see Sosa 1991), while the second reading is more patent in his presentation of the performance-based approach to knowledge (Sosa 2007). In more recent work (Sosa 2010, 2011), Sosa opts for the third reading. Although Turri certainly deserves credit for distinguishing these three readings as different ways of interpreting the 'because of' relation, the third reading, which he calls 'adeptness', already appears in Sosa's 2007 book, which is previous to Turri's paper (2011) (Turri acknowledges it). Consider this quote: "we might understand success due to an agent's competence as success that manifests that competence, a special case of the manifestation of a disposition" (Sosa 2007: 80). If Turri calls adeptness rather than aptness the third reading of Sosa's notion of aptness it is because he has the first reading in mind. To avoid complications, I will stick to Sosa's terminology and I will use the term 'aptness' to refer only to the condition stated

The motivation for this reading of the 'because of' relation comes from the debate on dispositions. A disposition (e.g., fragility), when triggered, may have many outcomes. However, only a subset of them may be said to *manifest* the disposition. What is meant by that? There is no explicit account in the epistemological literature and I do not think we need one. Turri (2011), for instance, treats 'manifest' as primitive and relies "on our robust pretheoretical understanding of it" (Turri 2011: 2). This pretheoretical understanding would amount to something like this: when an outcome manifests a disposition at the very least the outcome has not been caused (or determined) independently of the disposition. At any rate, the best way to grasp the manifestation relation is through examples. Turri (2011: 4) gives two illustrative ones:

> BOIL
>
> You place a cup of water in the microwave and press start. The magnetron generates microwaves that travel into the central compartment, penetrate the water, and excite its molecules. Soon the water boils.
>
> FIRE
>
> You place a cup of water in the microwave and press start. The magnetron generates microwaves that cause an insufficiently insulated wire in the control circuit to catch fire, which fire deactivates the magnetron and spreads to the central compartment. Soon the water boils.

According to Turri, the outcome in BOIL manifests the microwave's boiling power, but the outcome in FIRE does not. And he claims: "[w]e have a plain way to mark the distinction: in boil, but not fire, the microwave boils the water" (Turri 2011: 4).

Turri thinks (correctly) that in the same way as in FIRE the boiling of the water does not manifest the microwave's boiling power, in many Gettier-style cases the cognitive successes of the agents in question do not manifest their epistemic competences. For instance, in case F (the case of the snake in the tree considered above) Adam's cognitive success does not clearly *manifest* his (typically reliable) visual competence. After all, he mistakes a snake-looking branch for a snake.

In general, any case in which an agent exercises a globally reliable and cognitively integrated perceptual competence C and as a result mistakes an object O_1 for O_2 but for some reason forms a true belief about O_1 will be a case in which the agent's cognitive success fails to manifest C. For this very same reason, APTNESS rules out as a case of knowledge MARY, one of the cases of veritic fortune introduced in

above (APTNESS), i.e., to what Turri calls 'adeptness'. See Greco (2012) for further discussion on the different senses of 'because of'.

chapter 4: the agent, Mary, comes to believe truly that her husband is in the living room by looking at someone who looks like her husband.[34]

7.4.2.1 *The Situational View of Aptness*

As we have seen, FULL CREDITABILITY does not account correctly for cases of testimonial knowledge because it is too strong: in such cases the cognitive success of the recipient of testimony is not only explained by the exercise of her ability to select reliable informants but also by the the cognitive abilities of the producer of testimony. Is APTNESS too strong as well? The answer is negative and the reason is that APTNESS does not put a constraint on the minimal degree of salience that the agent's cognitive abilities must have in the explanation of her cognitive success in order for the agent to have knowledge. Let us see how APTNESS explains CHICAGO VISITOR.

The way Lackey describes the case makes one suppose that Morris is a competent speaker. Competent speakers have a generally reliable ability to select good informants such that, for each informant and each situation individually, the ability monitors (perhaps on the basis of stereotypes) those features of the informant and the situation that indicate that the informant is reliable (e.g., if one wants to know some directions, the ability prevents one from trusting children, people who look like tourists or a visibly crazy person uttering directions at random). In this way, it is plausible to think that Morris's cognitive success manifests his detection ability in *that* situation. This is consistent with the fact that Morris's cognitive success also manifests the informant's cognitive abilities, but more importantly, it is consistent with the fact that we ascribe knowledge to Morris in *that* situation.

Things are different in the following scenario. First, let be 'F' the set of features detected by a recipient of testimony that indicate her that her interlocutor is a reliable informant. Now consider the following case:

> ACTOR
>
> Morris, new in the city, wants to know the position of the Sears Tower. He looks for a good informant and selects Philip, who is dressed like a police officer, on the basis of F (F may include, for example, those features that make people trust the police in general). However, Philip happens to be an actor cleverly disguised as a police officer whose intention is to give wrong directions to outsiders (F indicates reliability but Philip is not reliable). It also happens that whenever Philip is about to lie about the landmark's location, a benevolent Genie moves the Sears

34 The same applies to STILL SNAKE, the case of veritic fortune presented in section 4.3 that is modeled on case F.

> Tower so that Philip speaks truly. Consequently Morris's beliefs about the position of the skyscraper are true.[35]

ACTOR is a case of veritic fortune. Unlike in CHICAGO VISITOR, here we do not have the intuition that Morris's cognitive success manifests his ability to select good informants. This does not mean that he does not have such an ability. He has it, and he exercises it, but his cognitive success does not manifest that exercise of ability. To give support to this intuition, note that ACTOR is structurally equivalent to cases of *veridical mimicking*, which are analyzed in the literature on dispositions.[36]

In cases of *mimicking* X does not have the dispositional property D to respond to stimulus S with response O, S obtains and something causes O. A paradigmatic example of mimicking is a rock with an explosive stuck such that if the rock were struck, it would shatter. The rock in this way 'mimics' fragile objects by behaving as a fragile object. In cases of *veridical mimicking* X does have the dispositional property D to respond to stimulus S with response O, S obtains and something extrinsic to X causes O. Here is an example by Johnston (1992: 233): a glass has the disposition to break when struck but a guardian angel has decided to break the glass when struck in a way that is independent of its fragility. In this way, the outcome (the breakage of the glass) *does not manifest* its fragility.

Let us compare veridical mimicking with ACTOR. In ACTOR, Morris [X] has a disposition to believe the truth through the testimony of informants that exhibit properties that indicate reliability [D]. Philip, the person who looks like a police officer, exhibits such properties [S]. On that basis, Morris accepts Philip's testimony and forms the true belief that the Sears Tower is located at certain position [O]. Due to

35 Curiously, ACTOR can be used to prove that Jennifer Lackey's recent definition of lying (Lackey 2013) is not sufficient for lying. Lackey's account can be reconstructed as follows: A lies to B if and only if (i) A states that p to B, (ii) A believes that p is false and (iii), in stating that p, A aims to conceal information from B or A aims to bring about a false belief in B. In ACTOR, (i) Philip states that the Sears Tower is at certain location L to Morris, (ii) Philip believes that it is false that the Sears Tower is at location L, (iii) Philip attempts to conceal information regarding the location of the Sears Tower from Morris (suppose that he has hidden all the evidence available in the surroundings that could show outsiders the location of Chicago landmarks) and he attempts to bring about in Morris the false belief that the Sears Tower is at location L. That is, conditions (i), (ii) and (iii) of Lackey's definition are satisfied. However, Philip's statement that the Sears Tower is at location L is not a lie in virtue of the following intuitive condition: A lies to B in stating that p, only if p is false. Therefore, the conditions of Lackey's definition are not jointly sufficient for lying. See Broncano-Berrocal (2013) for relevant discussion.

36 See Fara (2006) for an excellent overview on the huge literature on dispositions. In addition, see Henderson & Horgan (2009) and Brendel (2009) for interesting discussions on the relation between the epistemic notion of intellectual virtue and the metaphysical notion of disposition. See Gundersen (2003) for a defense of a so-called dispositional theory of knowledge and Gundersen (2010) for relevant discussion on the relation between problematic cases for the conditional analysis of dispositions and problematic cases for tracking analyses of knowledge.

the intervention of the Genie (i.e., the mimicker), that outcome (i.e., Morris's cognitive success) does not manifest Morris's disposition to believe the truth through reliable testimony. In a sense, the intervention of the Genie mimics Morris's disposition.

What is the difference between CHICAGO VISITOR and ACTOR? Is it a difference in competence such that Morris has the ability to select reliable informants in CHICAGO VISITOR but not in ACTOR? I do not think so. In both cases, Morris exercises his ability to detect reliable informants in the following way: in both situations he successfully identifies prototypical properties that indicate, in general, that a person is a reliable informant. Moreover, in ACTOR Morris selects his informant on a better basis than in CHICAGO VISITOR: while in the latter case Morris selects an informant on the basis that the person is an adult (rather than a child), in the former case he selects an informant on the basis that he is an adult and that he looks like a police officer (police officers and taxi drivers are typically the most reliable informants when it comes to giving directions).

Why does Morris acquire knowledge in CHICAGO VISITOR but not in ACTOR? Why does his cognitive success manifest his selection ability in one case but not in the other? The answer is simple: because the degree of competence that an agent must have in order for her success to manifest the relevant competence varies as a function of the environment (situation, context, circumstances). Morris's cognitive success in ACTOR would have manifested his detection ability if that ability had been very fine-grained and sensitive to minimal details that would have enabled him to detect that Philip was lying. However, although Morris is quite competent at selecting reliable informants, he is not so competent. Compare ACTOR with the following case:

> HUMAN LIE DETECTOR
>
> Having just arrived at the train station in Chicago, Morris, the famous professional poker player, wishes to obtain directions to the Sears Tower, where a poker tournament is taking place in a couple of hours. Morris has spent the last year studying cognitive science experiments on the prototypical facial expressions and gestures that people make when they lie (e.g., nose touching, mouth covering, sweating, fidgeting, saliva swallowing, and so on). He has become an excellent human lie detector. He looks around and approaches a person dressed like a police officer, Philip, who happens to be an actor whose intention is to give wrong directions to outsiders (he hates outsiders). Morris asks Philip how to get to his desired destination. As Philip starts talking, Morris starts noticing a lot of signs that indicate that Philip is lying. Morris po-

litely thanks Philip, says goodbye and goes away without believing a word of what he has just heard.

In Human Lie Detector, Morris's cognitive success (refraining from believing what he has just heard) certainly manifests his competence to detect lies.[37] The same cannot be said of Actor. Again: the degree of competence that one must have in order for one's success to manifest competence varies as a function of the environment. In Actor, Morris does not reach the level of competence that is required in order for testimony-based cognitive success to manifest competence. In Human Lie Detector, by contrast, the intensive training that Morris has carried out for the last year has increased his level of competence in detecting lies. This is why his cognitive success manifests his detection ability. Finally, in Chicago Visitor the level of competence required in order for a testimony-based cognitive success to manifest competence is much lower. This is why Morris comes to know the location of the Sears Tower with same level of detection ability that he displays in Actor, where he fails to know its location.

The phenomenon that I have just described is not circumscribed just to testimony cases. It is not even circumscribed to cases of *cognitive* performance. It applies to any type of performance. For example, the degree of archery competence that one must have in order for one's successful shot to manifest competence when shooting at one meter from the target is much lower than the degree of archery competence that one must have in order for one's successful shot to manifest competence when shooting in a windy night at one hundred meters from the target.

This variability is inherent to the concept of control and to the concept of achievement. In certain environments, one's successes do not manifest competence because in those scenarios the minimal level of competence required in order for that type of successes to manifest competence is higher than the maximum level of competence that one is able to exercise.[38] Those successes are neither under our control nor can be considered achievements. In some other cases, the level of skill required is much lower and our successes thus manifest competence. I call this the *situational view of aptness.*

Aptness must be interpreted accordingly, i.e., as putting no constraint on the degree of competence that an agent must have in order for her epistemic competences to be manifested by her cognitive success. The minimum level of epistemic competence required in each case will be a function of how the environment (the circumstances, the situation or the context) is. I do not think that there is a principled way to spell this out, because the relevant levels of competence

37 See Sosa (2011, ch. 1) for relevant discussion on how suspension of judgment counts as a cognitive performance with a distinctive aim: to avoid failure rather than to attain the truth.

38 Again, this does not mean that one is not competent in those situations; it only means that one is not *sufficiently* competent.

are set by pragmatic factors. Nevertheless, we can have an accurate grasp of which are the levels of competence required in each case by comparing cases of the same type (e.g., cases of testimony, perception, inference, and so on), as I have done with Chicago Visitor, Actor and Human Lie Detector. One might not find this method very convincing, but it is at least a way of testing our intuitions. And one will find it definitely more convincing than the treatment of the cases of the rival view of aptness, which I will present in the next section.

Before that, let me note that, unlike Full Creditability, Aptness does not rule out as cases of knowledge cases of knowledge by cognitively extended processes. The structure of those cases is analogous to the structure of cases of testimonial knowledge: something (a device) or someone (an informant) plays an important role in the explanation of the agent's cognitive success. Consider Sissicase. Is Sissi's true belief that the suitcase contains a bomb apt? In other words, does Sissi's cognitive success manifest competence? There is no reason to think otherwise. After all, although Sissi is more attentive thanks to System 2, Sissi exercises visual competence when she identifies dangerous objects. Therefore, her successful identifications manifest visual competence. In addition, her cognitive success also manifests the effects of working under System 2.

7.4.2.2 *The Appropriateness View of Aptness*

A common way of thinking about the solubility of salt is this: salt can only manifest its solubility under some 'normal' or 'appropriate' conditions, in particular only if it is introduced in a liquid with such-and-such characteristics. Analogously, some epistemologists (Kelp 2012; Sosa 2011) think that an agent's belief can only manifest cognitive competence only if it hits the truth under 'appropriate' conditions. These appropriate conditions encompass environmental conditions (e.g., good light conditions) and internal conditions of the agent (e.g., being sober).[39] Call the *appropriateness view of aptness* the view that a true belief can be assessed for aptness only if the conditions under which it has been formed are appropriate, i.e., the view that a true belief is apt only if the agent is appropriately situated and in appropriate condition.

We have seen that Aptness allows to explain why there is no knowledge in some cases of veritic fortune such as Mary or Still Snake. How is Aptness supposed to handle the trickier Temp? The way in which advocates of the appropriateness view would rule out this case as a case of knowledge would be to argue that the requirement for apt belief that the agent must be *appropriately situated* is not met. They would argue, for instance, that the unreliability of the thermometer makes the situation inappropriate, so that Temp's cognitive success

39 See Sosa (2010) for relevant discussion.

in that situation does not manifest his globally reliable competence to form true beliefs about the temperature.

Intuitive as it may be, appealing just to the inappropriateness of the conditions as a way of dealing with problematic cases should be the last resource of any theory of knowledge, as one may end up judging that the circumstances are not appropriate whenever a case looks like a counterexample to the theory. Let me illustrate the point with an example.

Consider a modification of Temp in which everything is as in the original case except for the fact that in an interval of time, T_1, the thermometer is broken and randomly fluctuating within a given range (as in the original case); in the next interval T_2 the thermometer works perfectly well; in T_3 the thermometer is broken and fluctuates again, and so on and so forth. If the intervals are long enough (e.g., one week long) we may presume that when Temp looks at the thermometer in an interval where the thermometer is working well, he knows the temperature of the room. If this is true, the advocates of the appropriateness view are compelled to say that the conditions are appropriate.

Now, if we start reducing the length of the intervals gradually, there is some point where we lose the intuition that Temp knows the temperature of the room. For example, if the intervals were one minute long, we would claim that Temp does not know the temperature (if the intuition does not arise, reduce the intervals, for example, to one second). The problem is that, according to the defenders of the appropriateness view, the circumstances during the one week long 'good' interval are appropriate because the thermometer works well but, at the same time, the circumstances during the one minute long 'good' interval are not appropriate *despite* the fact that the thermometer works well. In view of this, it seems that any explanation that the defenders of that view gave at that point of why the circumstances are inappropriate in the short interval case would be *ad hoc*.[40]

To be clear, I do not claim that an advocate of the appropriateness view cannot conclude that Temp's true beliefs about the temperature are inapt (in Temp or in the modified case). They can draw that conclusion. All I claim is that such a conclusion is *ad hoc* in the modified case. Let me compare this treatment of the cases with the way the situational view of aptness would explain them.

In my view, the degree of competence that an agent must have in order for her success to manifest competence (i.e., in order for her

40 For example, they will surely point to the fact that a thermometer that reliably indicates the temperature in the way described does not meet the standards of reliability required for aptness; alternatively, they would include information about the intervals in their judgments about whether the circumstances meet the standard of appropriateness. However, which length of time is enough to meet those standards? Whichever answer one gives should not assume that it is the length of time where we would ascribe knowledge to Temp, as that would be to beg the question. It does not seem that one can give an answer without making such an assumption, though.

successful performance to be apt) varies as a function of the environment. For each situation *S* there is a minimum level *L* of competence *C* that an agent possessing *C* must reach in order for her successful performance to manifest *C*. (*L* is determined by pragmatic factors and becomes evident when we compare cases of the same type). Note that there are two reasons an agent's successful performance might fail to manifest *C*: 1) the agent does not possess *C* (her performance is not adroit); 2) the agent possess *C* but her level of competence does not reach *L*. In addition, note that on this view there is nothing like 'appropriate' or 'normal' circumstances: all that matters is that one has the relevant competence and that one is able to exercise it with certain level of proficiency.

Suppose you have the competence in question. For example, suppose that you are a very competent archer. Plausibly, you would retain your competence in an environment where there is a magnet at the tip of your arrows and where the targets are magnets with same pole. Are those circumstances inappropriate? Advocates of the appropriateness view would claim: yes, they are. Suppose that one of your arrows ends somehow hitting the target. Should we think that your successful shot is not assessable for aptness because the circumstances are inappropriate?

In my view, your successful shot is assessable for aptness in the following way: your success does not manifest your archery competence (i.e., it is not apt) because the level of skill required in that situation in order for your performance to count as apt is much higher than the level of skill that you actually display. To reach such a level you would have to shoot with such a strength that you could make your arrows go through the magnetic fields. But you are not so strong and hence not so competent and hence your success does not manifest your archery competence. Aptness is therefore a gradual property that varies as a function of the level of skill set by the circumstances. Consequently, there is no need to split the circumstances into two groups: normal or appropriate on the one hand and abnormal or inappropriate on the other.

Let us go back to TEMP. We may assume that Temp is competent in trusting epistemic devices such as thermometer, watches or speedometers. What does it mean that one is competent in trusting this kind of devices? Basically, that the trust one places on their reliability is not arbitrary but grounded on an inductive basis or perhaps on a trial-and-error basis.[41] Imagine that Temp is in a situation in which there is no intervention of the hidden agent but in which the thermometer still fluctuates at random. If the range of fluctuation is very small, Temp will probably keep trusting the readings of the thermometer. However, if the range of fluctuation is wide (or very wide),

41 See Sosa 2011, ch. 7, for relevant discussion on instrumental knowledge (knowledge from epistemic devices).

Temp will not believe the readings (at least not all of them). For example, if the temperature in the room is around 20ºC and the thermometer reads '45ºC', Temp will surely not believe that the temperature is 45ºC, because he is able to detect that the temperature is not 45ºC. Such a cognitive success (avoiding failure) manifests competence and is, therefore, apt.

When the hidden agent intervenes, the level of competence that Temp needs in order for his cognitive success to be apt is extremely high. Such an intervention makes his competence to detect the temperature within certain margins and to trust thermometers practically useless. What would be required to achieve that level of competence? Temp's cognitive success would manifest competence if, for example, he were like Tempo, the agent with a temperature-detector implanted in his brain from birth. Other possibility would be that a reliable informant told Temp about the manipulation of the hidden agent. In such a case, his true beliefs about the temperature would manifest competence (namely his ability to detect reliable informants) and also the informant's competences.[42]

Consider now the 'interval' version of Temp. Unlike the appropriateness view of aptness, the situational view does not lead to the absurd conclusion that the circumstances during the one week long interval in which the thermometer works well are appropriate while the circumstances during the one minute long interval in which the thermometer also works well are not appropriate. The situational view accommodates the case in the following way. During the one week long interval the level of competence required for aptness is the usual one: one only needs a competence to trust devices properly (as I have described it above). Consistently, in that case we ascribe knowledge to Temp. By contrast, during the one minute long interval the required level of competence is much higher: in addition to trusting the thermometer properly one needs to discriminate *when* the thermometer is working well and when is not. Unless one is a thermometer technician, one will probably lack such a competence. Consistently, we do not ascribe knowledge to Temp in that case.[43]

Let me summarize the two views of aptness. The appropriateness view of aptness is inspired by the common way of thinking about

42 Sometimes, in situations that are very demanding from an epistemic point of view we can only achieve knowledge by competently relying on other people's cognitive abilities, or on reliable devices (e.g., another thermometer would have been a great aid for Temp). The same point can be made about the notion of control: in some situations you can only control *X* by competently relying on someone or something with causal influence over *X*. A CEO has control over the products of her company by competently relying on the competent people that produces them.

43 Presumably, there will be borderline cases in which it is not clear whether such a high level of competence is required for aptness (e.g., consider three hours long intervals) but that is also consistent with our hesitation to ascribe knowledge in those cases.

dispositions. In the same way as an entity (e.g., a glass) manifests its disposition (e.g., its fragility) under suitable conditions (e.g., when struck), a true belief manifests cognitive ability in appropriate circumstances. For example, the true belief that there is a red painting on the wall manifests visual competence when one looks at it under good light conditions, at proper distance, sober, sufficiently attentive, and so on. It is in this sense in which the belief is said to be apt: it is true because of competent under appropriate conditions. If the conditions are not appropriate, a true belief cannot be assessed for aptness (authors who endorse this view mean by this that the belief is not apt). As we have seen, the view has problems in cases in which it is not clear whether the conditions are appropriate or not. I have proposed a different way of thinking about aptness, one that steers clear of these problems. In my view, the degree of (cognitive) competence that an agent must have in order for her (cognitive) success to manifest competence (i.e., in order for her successful (cognitive) performance to be apt) varies as a function of the circumstances. Different circumstances require different levels of competence, where the minimum level of competence required is set by pragmatic factors. One might find this explanation lacking of specificity. However, this is how we commonly think about abilities, competences and skills: the degree of archery skill needed in order for a successful shot to manifest competence in an environment where the winds are calm is much lower than the degree of skill needed in order for a successful shot to manifest competence during a hurricane. The epistemic case is not different.

Pros and cons

We have seen that APTNESS (interpreted from the situational point of view) correctly rules out cases of intervening luck (e.g., case F) and cases of veritic fortune (e.g., MARY, STILL SNAKE, TEMP) as cases of knowledge. Unlike FULL CREDITABILITY, APTNESS is compatible with cases of knowledge from testimony and from cognitively extended processes.

However, like FULL CREDITABILITY, APTNESS is not a good anti-environmental-luck condition. In FAKE BARNS, for example, where Henry forms the true belief that the object in front of him is a barn by looking at a genuine barn, the causal relation between Henry's belief and what makes it true is the appropriate one: Henry's visual faculties are excellent, the light conditions are good, he is at few meters from the barn. In sum, everything is as good as in the case where he comes to know that the object in front of him is a barn. Therefore, we cannot but conclude that Henry's cognitive success does manifest his visual ability. To give support to this intuition, we can consider a couple of well-thought cases proposed by Christoph Kelp (2012: 9):

MINIMALIST ART

> Ernie is standing in front of a piece from a series of ten monochrome paintings that are currently being exhibited at the local art gallery and comes to believe that the canvas he is looking at is red.

Plausibly, Ernie knows that the painting he is looking at is red, so that advocates of Aptness would claim that Ernie's belief is apt. Consider now the second case:

> Conceptual Art
>
> Bert is standing in front of a piece from a series of ten monochrome paintings that are currently being exhibited at the local art gallery and comes to believe that the canvas he is looking at is red. Unbeknownst to Bert, he is looking at the only red monochrome in a series of otherwise white monochromes cleverly illuminated to look like red monochromes.

Plausibly, Bert does not know that the painting he is looking at is red. Could advocates of Aptness claim that Bert's belief is not apt? Defenders of the appropriateness view of aptness could claim so by appealing to the inappropriateness of the circumstances. However, in which sense Bert's beliefs are inappropriate? After all, the causal relation between Bert's belief and what makes it true is the appropriate one (e.g., suppose that Bert's visual faculties are excellent, that he looks at the canvass at one meter distance and that there is a strong white spotlight illuminating it). The conditions seem as appropriate as in Minimalist Art. Therefore, Bert's cognitive success manifests his visual abilities.

The same conclusion follows from the situational view of aptness. Suppose that in Conceptual Art Bert looks at the canvass at one meter distance and that there is a strong white spotlight illuminating it. Suppose that in Minimalist Art Ernie looks at the canvass at great distance and that it is a little dark. Suppose that Ernie knows that the canvas he is looking at is red and that, therefore, his cognitive success manifests competence. As the cases are described, it seems that the level of visual competence that Ernie must have in order for his belief to be apt is greater than the level of competence that Bert must have. Therefore, if Ernie's cognitive success manifests competence so does Bert's.[44]

44 One could argue that the presence of the cleverly illuminated white canvasses in Conceptual Art dramatically increases the level of competence required in order for a visual cognitive success to manifest competence. After all, one could argue, the exhibition is set up to deceive visitors. In this way, Bert's cognitive success would not manifest competence, i.e., it would not be apt, which would show that aptness is able to dispose of some cases of environmental luck. I am sympathetic to this reconstruction of the case as well. The problem, in any case, does not have to do, I think, with a failure to manifest a meta-competence and hence of the lack of meta-aptness. See Sosa (2011) for this kind of answer.

The initial question of this section was: How should we read the 'because of' in Success Because Of Competence? In terms of creditability or in terms of manifestation of competence? We can now give an answer: in terms of manifestation. Otherwise, the account would entail that one does not have control over one's cognitive performance (and hence that one lacks knowledge) in cases of testimonial knowledge and perhaps in cases of knowledge by cognitively extended processes. In this way, we can reformulate Success Because Of Competence as Aptness:

- Aptness: *S* controls her cognitive performance, only if *S*'s cognitive success (i.e., *S*'s believing truly that *p*) *manifests* epistemic competence (or cognitive ability).

We are now in a position to offer a more complete picture of the notion of epistemic control.

7.5 THE CONTROL THEORY OF KNOWLEDGE REVISITED

The control theory of knowledge says that knowledge is a cognitive achievement. A cognitive success is an achievement just in case the agent has control over her cognitive performance. What is meant by that? What is epistemic control? Let us put all the pieces together. In chapter 5, the conclusion was that:

- Safety: *S* controls her cognitive performance only if her cognitive success is safe.[45]

In this chapter we have seen that:

- Success: *S* controls her cognitive performance only if the resulting belief is true.

- Competence: *S* controls her cognitive performance only if she has the relevant cognitive competences or abilities.

- Cognitive Integration: *S* controls her cognitive performance only if *S*'s relevant cognitive abilities are cognitively integrated.

  (*S*'s cognitive disposition is cognitively integrated with other of *S*'s cognitive dispositions if and only if:

  – [Stability] it is sufficiently stable,

  – [Causal Interaction] it causally interacts with the other dispositions,

  – [Coherence] the beliefs it produces cohere with the beliefs produced by the other dispositions,

45 A true belief is *safe* just in case it meets the requirements of Safety I, II and III. See section 5.5.

- [Disposition to Coherence] the whole system of *S*'s cognitive dispositions is disposed to form and maintain a coherent set of beliefs,
- [Ownership (i)] *S* recognizes herself as the source of her beliefs,
- [Ownership (ii)] *S* accepts that she is fairly credited with having certain beliefs (the kind of beliefs that the disposition produces), and
- [Ownership (iii)] *S* bases her beliefs about herself on appropriate grounds [e.g., as a result of an appropriate learning process])

- Aptness (situational view): *S* controls her cognitive performance, only if *S*'s cognitive success (i.e., *S*'s believing truly that *p*) *manifests* epistemic competence (or cognitive ability).

In conclusion, epistemic control (and hence knowledge) requires both safety and aptness. In the next chapter, I will explain why the conjunction of safety and aptness does not seem to suffice for knowledge and I will evaluate several possible ways of overcoming this difficulty. In chapter 9, I will argue that safety and aptness are related in a very specific manner and that the notion of epistemic control must be understood in those terms. Before that, and for the sake of completeness, it is precise to discuss two further issues.

7.6 FURTHER ISSUES

7.6.1 *The Question of Voluntariness*

The most natural reaction when we put together the terms 'control' and 'belief formation' in a sentence is to think that there cannot be such a thing as controlling belief formation because there is nothing such as acquiring or withholding belief directly upon performing an act of the will. Many will think that since 1) control is voluntary and 2) believing is not voluntary, believing cannot be controlled. This argument has been blocked by Pamela Hieronymi, whose excellent work on attitudes (2006, 2008, 2009) shows, among other things, that the idea of involuntary control (and agency) over belief makes perfect sense. In what follows, I will present those parts of Hieronymi's work that will help the control theory of knowledge to steer clear of the question of voluntariness.

7.6.1.1 *Hieronymi on Voluntariness and Belief*

Hieronymi's account of attitudes hinges on a central assumption: certain attitudes (e.g., belief and intention) embody their subject's answer to some question or set of questions (Hieronymi 2009). For instance, by settling positively for oneself the question of whether to φ (e.g., buy a new laptop), one therein intends to φ. Furthermore, if one intends to φ, she argues, one is *committed* to φ-ing, in the sense that having such an intention makes one vulnerable to criticisms and questions of the type 'Why do you intend to φ?', which one can only answer correctly by appealing to the reasons one takes to bear positively on the question of whether to φ.[46] Hieronymi calls attitudes with this structure *commitment-constituted* attitudes.

Crucially, believing has the same structure according to Hieronymi: by settling positively for oneself the question of whether *p*, one therein believes that *p*. Furthermore:

> [I]nsofar as one believes *p*, one is committed to a positive answer to the question of whether *p*, i.e., one is vulnerable to a range of questions and criticisms that would be satisfied by reasons that (one takes to) bear positively on whether *p*. So we can say that a belief that *p* embodies a positive answer to the question of whether *p*. (Hieronymi 2009: 139).

§ *Voluntariness*. Let us move on to the question of voluntariness. What does it mean that an activity is voluntary?

> [A]n activity is voluntary just in case you decide to do it for reasons you take to settle the question of whether to do it, therein intend to do it, and, providing all goes well, do it by executing that intention. (Hieronymi 2008: 366)

Buying a new laptop is obviously voluntary on Hieronymi's account. If after having considered the pros and cons of buying a new laptop one takes certain set of reasons to settle positively the question of whether to buy it, one therein intends to buy it, and one subsequently executes the intention (if one is able to). Unlike buying a laptop, believing is not voluntary. Hieronymi is categorical on this: one cannot believe *p* by settling the question of whether to believe *p*; one can only believe *p* by settling the question of whether *p*.[47] For example, some scientific research seems to show that believing in God is beneficial for health because it leads to a more contended life.[48] If you

46 For example, imagine that your laptop stopped working this morning and you urgently need a new one to finish a paper that you must submit tomorrow. If you take this reason to settle positively the question of whether to buy a new laptop, you therein form the intention to buy it, and if someone ask you why do you intend to buy a new laptop, you will answer correctly by appealing to this reason.

47 Hieronymi calls 'constitutive reasons' for the belief that *p* reasons taken to bear positively on whether *p* and 'extrinsic reasons' reasons taken to bear positively on whether to believe that *p* (i.e., on whether it is good to have such a belief). See Hieronymi (2006).

48 See Clark & Lelkes (2009).

are a convinced atheist you may find perfectly reasonable that the belief in God is a belief that it is worth having (because of its benefits for health) without thereby coming to believe in God. Another example: a xenophobic politician may come to think that it would be very useful to believe, during political campaign periods, that immigrants should have equal rights (so she can avoid lying all the time) without thereby coming to form such a belief. All the politician would form is a second-order belief about the eventual belief that immigrants should have equal rights.

§ *Are control and agency always voluntary?* If believing is not voluntary but control is, is therefore belief beyond our control? Let us pause here for a moment. Is control *always* voluntary? Why we tend to think that control is always voluntary? Hieronymi explains that we have that impression because agency and control are typically exercised in ordinary intentional action: paradigmatic exercises of agency or control consist in someone executing one's intentions with respect to an action or object (Hieronymi 2009: 144). This kind of exercise of agency or control has, according to Hieronymi, two distinctive features: *voluntariness* and some sort of *reflective distance* or *awareness* (e.g., of the intention in question).

However, Hieronymi thinks that there is a pervasive form of control that does not share these features:

> [I]f you become convinced that *p*, and so settle for yourself the question of whether *p*, you therein, *ipso facto*, believe *p*. Likewise, if you settle (positively) the question of whether to φ, you therein, *ipso facto*, intend to φ. Moreover, if you change your mind about whether to φ, or about whether *p*, in such a way that you are no longer committed to φ-ing or to the truth of *p*, then you no longer intend to φ or believe that *p*. We might say that we control these aspects of our minds because, as we change our mind, our mind changes—as we form or revise our take on things, we form or revise our attitudes. I call this exercising *evaluative control* over the attitude. (Hieronymi 2009: 139-140)

She asks: why should evaluative control deserve to be considered a kind of agency? She makes two points. For the one thing, it would be a mistake to put believing and intending in the same category as digesting, sneezing or blinking. All of them are certainly involuntary but, unlike the latter, the former are not just things that befall to one, in the sense that one is not passive with respect to them. For another, if forming and revising our intentions is neither voluntary nor something of which we are aware (because we intend only for reasons that settle the question of whether to act and not for second-order or deliberative reasons about whether is good to act), then voluntariness and awareness are not necessary features for agency. Why? As Hieronymi points out, precisely "because it seems that the forming of an intention must be an exercise of agency, if anything is" (Hieronymi 2009: 150).

7.6.1.2 *Three More Varieties of Control*

§ *Voluntary, evaluative and merely functional control*. I am very sympathetic to the idea that there is an intermediate category between the *voluntary control* we have when we perform intentional actions and the *sub-personal* or *merely functional* control that the nervous system has over basic biological processes such as digestion or blood circulation.[49] In this way, the controlled objects of this intermediate category are neither voluntary activities (e.g., kicking a ball) nor involuntary processes that befall to us and that we can only influence through our voluntary actions, that is, indirectly (e.g., you can take an anticoagulant to facilitate blood circulation). In this intermediate category, we find processes such as formation, revision and maintenance of belief, which are neither voluntary nor things that befall to us, i.e., with respect to which we are passive. According to Hieronymi, an important feature of *evaluative control* is the following:

> There is no reflective distance in the offing. We change our minds, and so control our attitudes, not by reflecting on or thinking about our mind, but rather by thinking about the object of our thoughts. The controlling happens "behind the lens," so to speak. The thinking subject controls its thoughts in thinking them. (Hieronymi 2008: 370-1)

A soccer player kicks the balls and the ball goes where the soccer player wants. We say that she has (voluntary) control over the ball. Unlike the soccer player, who might reflect on the desired trajectory of the ball before kicking it, one does not change one's belief that p by reflecting about whether it is a desirable mental state to be in, but by reflecting on whether p is true or not. In doing so we can say that one has (evaluative) control over one's belief. Evaluative control, in this way, is a very special form of control.

§ *Tracking and effective control*. These three varieties of control are compatible with the notions of tracking and effective control. Let me give some examples. There is voluntary tracking control: the kind of control that a doctor has over an infection when she runs tests on her patient to determine its cause. There is also voluntary effective control: the kind of control that a surgeon has over her scalpel when she operates on a patient. There is merely functional tracking control: the kind of control that a thermostat has over the temperature of the environment when the heating is off and the sun keeps the temperature

49 Hieronymi does not distinguish these varieties of control, but the distinctions follow from her account.

constant at the desired value (the same applies to thermoreceptors, neurons that are sensitive to changes in temperature). And there is also merely functional effective control: the kind of control of a thermostat when it activates or deactivates the heating (the same applies to thermoregulatory processes like perspiration). How can we accommodate the notions of tracking and effective control in the case of evaluative control?

Evaluative control over belief arises when one considers whether a proposition is true or false and as a consequence one starts, maintains or stops belief in that proposition. This process is for the most part automatic and is meant to include cases of automatic perceptual belief, but it can also be reflective, as when one considers whether one is speaking too loud when giving a speech. One might consider whether a proposition is true or false in a competent or in an incompetent manner, that is, one's consideration might arise from an exercise of cognitive ability or not. A case of incompetent evaluative control would be, for example, a case in which someone changes her mind after hearing a blatant lie or a case in which a myopic person comes to believe that there is someone over there by mistaking a coat hung on the coat hanger for a person. A case of competent evaluative control would be, for example, a case in which someone changes her mind after carefully investigating the veracity of a report or a case in which someone with 20/20 visual acuity comes to believe that there is someone over there after seeing a person over there.

We may say that an agent *S* has *effective evaluative control* over the belief that *p* if and only if the fact that *S* comes to believe the truth about *p* is because of *S*'s cognitive abilities (aptness).[50] Effective evaluative control is the kind of involuntary doxastic control required for knowledge. However, knowledge is a special state of mind, one that requires, in addition, another special type of evaluative control. To see this, let me illustrate again how Hieronymi conceives evaluative control over belief:

> Because [beliefs] embody our take on the world, on what is or is not true (...), we control them by thinking about the world, about what is or is not true (...). Because our minds change as our take on the world changes—because our minds change as we change our minds—we can be said to be "in control" of our [beliefs]. (Hieronymi 2008: 370-1)

When one changes one's take on the world and comes to believe something true, one's doxastic state might also be true in a range of close counterfactual situations in which one would come to believe the same proposition in the same manner. When that happens we may say that one's epistemic evaluation (and the resulting doxastic state)

50 The same definition applies to belief maintenance and disbelief.

tracks the truth (safety). It is in this sense in which we may say that an agent has *tracking evaluative control* over one's belief.[51] Only when one has both tracking and effective evaluative control over one's beliefs one's cognitive success may be said to be a complete cognitive achievement. Knowledge, as we have seen, is a complete cognitive achievement.

7.6.2 *Lottery Cases and Epistemic Control*

In chapter 4, when I introduced the hypothesis that nearly all (if not all) cases of true belief that is not knowledge are cases either of veritic luck or of veritic fortune (Hypothesis), I anticipated that it could be difficult to explain why some cases of not-known true belief are cases of veritic fortune. In particular, the cases I mentioned there were cases of reflective luck that do not involve skeptical hypothesis (e.g., cases of misled chicken-sexer cases) and lottery cases. We have already seen how to accommodate the former.[52] It is time now to explain how to accommodate the latter.

Here is the puzzle. Most philosophers think that, if you buy a lottery ticket, and you come to believe that your ticket is not the winner in virtue of the high probability that it had of not being the winner, you do not know that your ticket is not the winner.[53] Suppose that your belief is true, that your ticket is not the winner. We have a case of true belief that is not knowledge. Is your belief luckily true? No, it is not: given the high probability of losing prior to the lottery draw, there was no risk that that your true belief that your ticket is not the winner could not easily have failed (or so it seems). If Hypothesis is correct, then your belief should be an instance of veritic fortune given that it is not knowledge. The question is: what makes it veritically fortunate? Why should we think that you lack epistemic control in lottery cases? Which sense of epistemic control fails: effective evaluative control (aptness) or tracking evaluative control (safety)? In what follows, I will discuss both possibilities: the possibility that the belief that one's ticket is not the winner is not apt and the possibility that it is not safe.

---

51 A similar point can be made about belief maintenance: when one keeps believing something true, one's doxastic state might match the truth in a range of counterfactual scenarios in which one would keep believing the same proposition in the same manner. On the other hand, tracking evaluative control also arises when in coming to believe a proposition, one would not believe closely related false propositions, propositions with gappy content, paradoxical content or with no content at all in close possible worlds.

52 See fn. 10.

53 See Hawthorne (2004) for an excellent monograph on the topic.

7.6.2.1 *First Possible Solution: No Aptness*

To show that the belief that one's ticket is not the winner is not apt, I will resort to two cases proposed by Jonathan Adler (2005). Adler asks us to imagine a company that manufactures widgets. All workers know that the manufacturing system produces one defective widget out of every thousand. In order to reduce the rate of defective widgets, the company managers have developed a detector that reads 'OK' when the widget is not defective and 'Defect' when defective. Adler asks us now to imagine two workers: Smith and Jones. Smith never uses the detector; instead, he plays video games all the day and puts an 'OK' stamp on every widget that comes to his station. After all, he thinks, he knows we will be right 99,9% of the times. Jones, by contrast, diligently applies the detector to each widget, carefully and skilfully (it is not an easy task to use the detector). The detector is accurate 99, 7% of the times (for some unknown reason, perhaps environmental distortion, the detector judges incorrectly three widgets out of every thousand).

Intuitively, Jones can come to know that a widget she tests is OK, and her cognitive success certainly manifest her competence to use the detector. By contrast, Smith does not know that a widget he tests is OK and intuitively his true beliefs do not manifest competence. Why do Smith's true beliefs fail to manifest competence?

As Adler describes it, Smith's method of belief-formation would be something like "assigning an object (e.g., a widget) to a category (e.g., 'not defective') on the basis of objects of that kind belonging to the category in an extremely high proportion of cases" (Adler 2005: 453). The problem with this method is that the inclusion of a widget in the category 'not defective' is carried out merely in virtue of the fact that it is a widget and independently of the features that make such a widget defective or non defective. As Adler points out, the method, so described, guarantees failure, and it is on the basis that each case is judged. This is why we do not have the intuition that Smith's true beliefs manifest competence. By contrast, Jones's true beliefs do manifest competence, because the detector tracks the features that make a widget defective and, therefore, the inclusion of a certain widget in the category 'defective' is not done independently of those properties but in virtue of them.

Analogous considerations apply to the lottery case. In the lottery case, you form your belief that your ticket is not the winner by assigning the ticket to a category ('loser') on the following basis: all tickets of the lottery except one are losers. Your inclusion of the ticket in this category is done, therefore, independently of the features that make your ticket a loser, i.e., independently of the fact that the balls that would make your ticket win are not the ones that actually come out (and just in virtue of the fact that your ticket is a ticket of the lottery). This is why your true belief is not apt.[54]

By contrast, if you *see* which balls come out (or if you read the result a newspaper) you can aptly come to know that your ticket is a loser. In those cases, your cognitive success manifest epistemic competence. To compare, consider cases of knowledge by inductive reasoning. When one comes to know a proposition of the form <*a* is a *F*> by inductive reasoning one's reasoning is based on a past record of regularities in which items similar to *a* are *F*'s. That one comes to believe the truth is explained by the fact that one's reasoning takes into account those regularities: given them, the inclusion of *a* in the category '*G*' is not done independently of the properties that make *a* a *G* (unlike in the lottery case), but in virtue of one's knowledge that most *a*'s have been *G*'s in the past.

7.6.2.2 *Second Possible Solution: No Safety*

One might still have the intuition that the belief that one's ticket is not the winner *does* manifest competence (in particular, probabilistic competence). Those who have that intuition might agree with the treatment that Duncan Pritchard (e.g., 2012c) makes of lottery cases. As we have seen in part I, Pritchard thinks that the extent to which an event is lucky is essentially a function of the closeness of possible worlds in which the event would not occur (provided that the relevant initial conditions are the same as in the actual world). But he also thinks that the closer it is a possible world in which an event would not occur, the luckier it is the occurrence of that event in the actual world.

If we put the latter idea in epistemic terms, this would mean that the closer it is a possible world in which one's belief is false, the luckier it is that one comes to believe the truth in the actual world. In the lottery case, one comes to believe truly that one's ticket is not the winner by probabilistic calculation. Given the high probability of

54 One could argue that there is a disanalogy between Smith's case and the lottery case: while Smith does not manifest all the competence that he could manifest given his epistemic situation (he could use the detector), in the lottery case one manifests all the relevant competence that one could manifest. However, this might not be true because one could perhaps buy a newspaper in order to come to know the result of the lottery. In any case, even if one could not do anything like that, the point is not that Smith's belief is not apt because he could have manifested more competence than he actually does and that the lottery player's belief that her ticket is not the winner is apt because she manifests all the competence that she could have manifested given her epistemic position. The point is that given their *actual* methods of belief formation (assigning an object (a widget/a ticket) to a category ('not defective'/'not winner') on the basis of objects of that kind belonging to the category in an extremely high proportion of cases) the competence that they *actually* manifest does not suffice to reach the level of competence that is required for their cognitive successes to manifest competence in their situations. In other words, their beliefs are not apt (according to the situational view of aptness).

losing, one's belief would be true in most close possible worlds (and presumably safe). However, Pritchard argues that in the *very closest* possible worlds, one's belief might not be true (because they are very similar and a slight configuration in the falling of the balls could easily produce any result).

Pritchard stipulates then an account of our tolerance to risk of epistemic failure. He thinks that the further a close possible world is, the more tolerant we are to epistemic failure. However, when it comes to the very closest possible worlds we are extremely intolerant to epistemic failure: it suffices a couple of possible worlds in which one's belief is false to make one's belief true by luck in the actual world (and hence to make it unsafe). This is actually what happens in the lottery case: one's true belief that one's ticket is not the winner would be thus unsafe.

I tend to agree with the idea that the belief that one's ticket is not the winner is not apt (because of the existing dissimilarity between lottery cases and, for instance, cases of inductive knowledge), but Pritchard's solution is valid as well. Perhaps our lack of epistemic control in lottery cases is explained both by a lack of aptness and a lack of safety. I leave the question open.

7.7 SUMMARY

This chapter has been a long exploration of the notion of epistemic control. The aim was to find an epistemic analogue of the notion of competent effective control. In section 7.1, 7.2 and 7.3, I have argued that epistemic control entails 1) cognitive success, 2) competent cognitive performance and the integration of the agent's cognitive dispositions. I have extensively discussed the notion of cognitive integration and I have provided a definition. In the most important section, section 7.4, I have argued that epistemic control entails that the agent's cognitive success is because of cognitive ability. I have devoted extensive discussion to the 'because of' relation. On the one hand, I have rejected a reading of 'because of' in terms of explanatory salience or creditability. On the other hand, I have argued that 'because of' should be read in terms of manifestation of competence. In particular, I have proposed a very particular way of understanding this reading: the extent to which an agent's (cognitive) success is because of (cognitive) ability depends on whether the agent reaches the level of ability that pragmatic factors of her present situation require in order for a (cognitive) success to be because of (cognitive) ability in the circumstances. In section 7.5, I have put all the pieces together and I have provided a comprehensive picture of the notion of epistemic control. In section 7.6, I have tackled two problems: how can it be that one has involuntary control over belief and how can it be that one lacks epistemic control in lottery cases.

8 SAFETY *BECAUSE OF* COGNITIVE ABILITY

I have defined knowledge as a complete cognitive achievement and complete cognitive achievement as a cognitive success (a true belief) that results from a cognitive performance over which the agent has epistemic control. Recall that a complete achievement is a success over which the agent has both tracking and competent effective control and that the safety and aptness conditions have allowed us to translate these two types of control into epistemic terms. In chapter 5, safety (a truth-tracking condition) was presented as the epistemic counterpart of the notion of tracking control (monitoring). In chapter 7, competent effective control has been cashed out in terms of aptness.

Tracking and effective control are typically not independent of each other. Think about a captain of a sailboat who has set sail heading South and who wants to hold that course. She notices that the winds are changing direction so she switches the sail from one side of the boat to the other in order to hold course. The captain's performance is successful, competent and also safe: not easily would the boat change course given the actual display of sailing competence (viz., of the competence to monitor the direction of the wind and the competence to adjust the sails accordingly). In a sense, the safety of the boat's course is *because of* the captain's sailing competences.

Analogously, the safety of an agent's cognitive success might be *because of* the agent's epistemic competences. In this chapter, I will analyze two accounts of knowledge that explore this link. Before that, in section 8.1, I will explain why the mere conjunction of safety and aptness does not seem to be sufficient for knowledge. In particular, I will present a couple of counterexamples to the sufficiency for knowledge of the so-called safe-apt view, a view recently proposed by Christoph Kelp (2012) that holds, precisely, that knowledge is safe, apt belief.

In section 8.2, I will analyze Duncan Pritchard's anti-luck virtue epistemology (Pritchard *et al.* 2010, 2012b, 2012c), a view that combines the safety condition with Partial Creditability. The way Pritchard formulates his account is compatible with two interpretations: knowledge as the conjunction of safety and Partial Creditability and knowledge as safe belief whose safety is partially creditable to cognitive ability. Although the former reading is the one that Pritchard seems to have in mind, I will also pay attention to the latter because it specifies in explanatory/creditability terms the idea that knowledge is safe belief whose safety is because of cognitive ability. I will argue that anti-luck virtue epistemology, in either interpretation, is not sufficient for knowledge

Section 8.3 will be a good opportunity to state more generally the kind of epistemological problem that underlies the counterexamples to Pritchard's anti-luck virtue epistemology. I will introduce the *direction of fit problem* (the problem of explaining under what conditions it is guaranteed the direction of fit that known beliefs have). I will argue that only the situational view of APTNESS provides a satisfactory solution to the problem.

Finally, in section 8.4 I will briefly evaluate John Turri's recent proposal of knowledge as ample belief (Turri 2011; *forthcoming*). Ample beliefs are beliefs whose safety manifests epistemic competence, i.e., amplitude is a specification in terms of manifestation of competence of the idea that known beliefs are beliefs whose safety is because of competence. Although I am quite sympathetic to Turri's view (I will defend a similar (although much more specific) account of knowledge in chapter 9), I will state some concerns regarding its lack of specificity and its motivation.

8.1 THE SAFE-APT VIEW

Christoph Kelp (2012) has recently offered a novel virtue-theoretic account of knowledge that he calls the *safe-apt view*:

- SAFE-APT: *S* knows that *p* if and only if *S*'s belief that *p* is safe *and* apt.

Kelp is a defender of the appropriateness view of aptness, which as we have seen is problematic because it explains certain cases in an *ad hoc* manner.[1] But he could well endorse the situational view, which avoids such complications. On the other hand, Kelp does not opt for any particular version of the safety condition. The only requirement that he mentions is the standard clause that safety does not require avoidance of error in all close possible worlds but only in those in which the agent acquires her belief via the same method that she uses in the actual world. Interestingly, Kelp's preferred way of individuating methods connects the two epistemic conditions of his theory. According to Kelp (2012: 12, fn. 11):

- Two method tokens m_1 and m_2 are of the same type *M* just in case they correspond to the same inner competence of certain cognitive competence.

§ *Inner competence*. Let me explain what is meant by 'inner competence'. Kelp follows Sosa (2010) in thinking that *S* has competence *C* only if:

1 See section 7.4.2. I will come to this point again in section 8.3.2.

- (i) *S* is physically and psychologically constituted in the way required for having *C* (e.g., the relevant constitution for a visual competence may include, among other things, the rods and cones of the human eye),
- (ii) *S* is in appropriate internal condition (e.g., in good shape, sober, awake, sufficiently attentive, and so on), and
- (iii) *S* is appropriately situated, i.e., in circumstances suitable to deploy *C* in a sufficiently truth-conducive manner.

In Sosa's terminology, (i) and (ii) are the inner component of *C*, viz., *S*'s inner competence. Before explaining why safety and aptness do not suffice for knowledge, let me present an objection to this way of individuating methods of belief formation.

8.1.1 *Puzzling Individuation of Methods*

Kelp accepts that one manifests visual competence when one looks at an object and forms a belief about that object even though one could easily have mistaken replicas of that object for the object. For example, he accepts that in FAKE BARNS Henry manifests visual competence when he looks at a genuine barn in the Fake Barn Country and therefore that his belief that there is a barn before his eyes is apt. On the other hand, the standard diagnosis of the case, which Kelp shares, tells us that Henry does not know (because in most close possible worlds in which Henry forms the belief that there is a barn before his eyes using the same method of belief formation that he uses in the actual world that belief is false).

To secure this verdict, Kelp needs Henry's belief to be unsafe, and to reach that result, it is essential that the belief-forming method used by Henry in the actual world is the same as the method that he uses in close possible worlds in which he looks at fake barns. Otherwise, the safety condition would hold and SAFE-APT could not rule out the case as a case of knowledge. Given Kelp's preferred way of individuating belief-forming methods, this means that Henry must have the *same inner competence* in the actual and in close possible worlds, which means, in turn, that his constitution and internal condition must be the same in all the relevant worlds. Consider now this other case of object identification *and* knowledge:

> DACHSHUND
>
> Oscar is standing in an open field containing Dack the dachshund. Oscar sees Dack and forms the belief that the object over there is a dog. Further suppose that Oscar has a tendency to mistake wolves for dogs (he confuses them with Alaskan malamutes, or German shepherds). If the object Oscar saw were Wiley the wolf, rather than Dack the

> dachshund, Oscar would (still) believe that the object over there is a dog. But surely it is wrong to deny that Oscar knows that proposition when he sees Dack the dachshund. The mere fact that he would erroneously take a wolf to be a dog hardly shows that he doesn't know a dachshund to be a dog! (Adapted from Goldman 1976: 779)

I take Goldman's diagnosis of the case to be correct: Oscar knows that the object over there is a dog when he sees Dack the dachshund even though he could *easily* have looked at Wiley the wolf. In the light of Kelp's treatment of FAKE BARNS, there is no reason why Oscar's belief that there is a dog over there when he looks directly at Dack should not count as apt. Crucially, Kelp needs Oscar's belief to be safe to secure the verdict that Oscar's belief amounts to knowledge,[2] and to reach that result, he needs to interpret the case in such a way that Oscar's actual visual method is *different* to the method that he uses in close possible worlds in which he mistakes Wiley the wolf for a dog (presumably, most close possible worlds). Can he achieve that result?

According to Kelp, two method tokens m_1 and m_2 are of the same type M just in case the inner components (the internal condition and the physical and the psychological constitution) of the relevant competences are the same. Arguably, the internal condition of Oscar's visual competence is the same in all the relevant worlds, since he is awake, sober, sufficiently attentive, and so on. Therefore, what must be different is the constitution component of Oscar's visual competence in close possible worlds in which he looks at Wiley the wolf.

However, why should we accept that Henry (in FAKE BARNS) has the required physical and psychological constitution for object identification when he looks at fake barns in close possible worlds but Oscar lacks such a constitution when he looks at Wiley the wolf? After all, both Oscar and Henry possess functioning and accurate visual systems (they have healthy retinas, optic nerves, visual cortices, and so on). In addition, they respectively possess the relevant concepts of barn and dog, they know the extension of these concepts (e.g., they know that wolfs do not fall under the concept of dog and that barn façades do not fall under the concept of barn), and so on.

Therefore, *either* both Henry and Oscar enjoy the required constitutional bases in close possible worlds where they misidentify objects *or* both agents are not constituted in that way in close possible worlds. Note that any satisfactory reply to this objection must be principled, in the sense that it must rely on a principle of individuation of competence internal components that allows to offer an explanation that can be generalized to other cases (for the same reason, the safety theorist

2 In particular he needs that the following holds: that in most close possible worlds in which Oscar acquires his belief that the object over there is a dog *via the same method of belief formation that he uses in the actual world* Oscar's belief is true.

needs a principled way of individuating methods).[3] In the absence of such a principle, SAFE-APT fails to explain the cases in a satisfactory manner.

8.1.2 *Safe, Apt, Not-Known Belief*

The following case might be considered a counterexample to SAFE-APT:

> FAKE BARNS II
>
> Henry looks at the only genuine barn in the field and non-inferentially forms the true belief that some object in the field is a barn (or the belief that there is a barn in the field).

Henry's belief is apt: in the actual world, he comes to believe the truth because of his visual abilities. Henry's true belief is safe: in close possible worlds in which he mistakes a barn replica for a barn there is still a barn in the field (namely the one that he is actually seeing). However, Henry's belief does not (intuitively) amount to knowledge. Therefore, Henry's belief that some object in the field is a barn is safe, apt and yet not knowledge.

§ *Possible reply.* Henry's belief is unsafe because in FAKE BARNS II there is risk of believing a closely related false proposition (BELIEF-FOCUSED RISK II), for example, the proposition that the barn he is looking at is a barn (when looking at a fake barn).

§ *Rejoinder.* In the same way as we did with STILL SNAKE in chapter 4,[4] the risk of believing a closely related false proposition can be excluded from FAKE BARNS II by modifying the case in such a way that the only belief that Henry could have formed in possible worlds in which he looks at fake barns is the belief that some object in the field is a barn. Suppose for example that if Henry had been about to form any other belief when turning his eyes to a fake barn, he would have been killed (imagine whatever details make the killing intuitively plausible for you). In this way, there is no longer a close possibility of believing a closely related false proposition.

§ *Possible reply.* Another possible line of response would be the following: the warrant for Henry's belief that some object in the field is a barn depends solely on the warrant for the demonstrative belief that the object before his eyes is a barn. The fact that the warrant for the demonstrative belief is defective (the belief is unsafe) makes the warrant for the belief that some object in the field is a barn also defective. Of course, we would need to spell out the nature of the dependence

3 I will provide such a principle in the next chapter.
4 See section 4.3, fn. 29.

relation between the demonstrative belief and the belief that some object in the field is a barn. However, this could in principle explain why the latter is not knowledge.

§ *Rejoinder.* If the lack of safety is the reason that the warrant for the demonstrative belief that the object before Henry's eyes is a barn is defective, we can modify the case as before, i.e., by blocking the possibility that Henry forms beliefs different from the belief that some object in the field is a barn when looking at fake barns. In this way, the actual demonstrative belief that the object before his eyes is a barn would not be unsafe (there would be no close possible world in which Henry would believe the same proposition falsely).

However, some might find this move not very compelling. What if the demonstrative belief is really unsafe? Would that mean that Henry's actual true apt belief that some object in the field is a barn is also unsafe (or epistemically defective in some other sense)? Perhaps. However, the defective warrant of the demonstrative belief might not be transmitted to Henry's belief that some object in the field if he forms the latter *inferentially*. Consider the following case:

> FAKE BARNS III
>
> Henry looks at the only genuine barn in the field and forms the true belief that the object in front of him is a barn. From that belief, he competently deduces and thereby comes to believe that some object in the field is a barn (or the belief that there is a barn in the field).

Henry's belief seems apt because he competently deduces it from another apt belief (the belief that the object in front of him is a barn). It also seems safe for the reasons exposed before. However, Henry's belief is not knowledge. The case then is a counterexample to the sufficiency for knowledge of SAFE-APT. The question is: is the second possible line of response to FAKE BARNS II also applicable (after proper modifications) to FAKE BARNS III? It is. Let us see how.

§ *Possible reply.* Here again, one could argue that since the warrant for Henry's belief depends on the warrant for the demonstrative belief that the object before his eyes is a barn, and Henry's demonstrative belief is unsafe, Henry's belief that the some object in the field is a barn is also unsafe because the premise belief (the demonstrative belief) is not safe and for some reason inference transmits lack of safety in this particular case.

§ *Rejoinder.* However, how can it be that Henry's competent inference from an unsafe apt belief does not transmit knowledge to a safe apt belief? The difficulty of giving a satisfactory answer precisely stems from the fact that Henry *competently* makes an inference from a *competently* formed belief. Is, therefore, the *lack of safety* of the premise belief what prevents knowledge of the conclusion?

I do not think it is. A powerful reason is that there does not seem to be any correct principle of (lack of) safety (or knowledge) transmission to which one can resort to justify with some degree of universality the point that the premise belief does not yield knowledge in Fake Barns III. For example, the following principle is not correct: if *S* does not know that *p*, competently deduces *q* from *p*, and believes that *q*, then *S* does not know or come to know that *q* on the basis of that deduction. Here is a counterexample by Peter Baumann (2012):

> Singing
>
> *S* forms the true belief that *A* is singing in the other room. However, it is usually *B* that does the singing around there and *S* cannot distinguish *A*'s from *B*'s voice, so that *S*'s belief is luckily true. However, if *S* infers from that belief that someone is singing in the other room, then the belief in that conclusion amounts to knowledge. (Adapted from Baumann 2012: 11)

Singing parallels Fake Barns III in the relevant aspects. First, the agents in question identify certain items (a barn, *A*'s voice) by visual or auditory competence even though they could easily have mistaken similar items (a fake barn, *B*'s voice) for the items they successfully identify in the actual world. Second, the propositions believed as a result of that successful identification are the premises of certain inferences, which result in new beliefs with the content <there is an item of type *T* at location *L*>. Third, the beliefs that result from those inferences are true. Now, if the structure of Fake Barns III is the same as the structure of Singing, why is there knowledge in the latter case but not in the former? In the absence of a satisfactory explanation, there is no other option but to conclude that Kelp's safe-apt view is insufficient for knowledge.

8.2 anti-luck virtue epistemology (alve)

Duncan Pritchard (*et al.* 2010, 2012b, 2012c) has recently shifted his view from an account that primarily understands knowledge in terms of the safety principle (a view mainly developed in his 2005 monograph on epistemic luck) to a variety of virtue epistemology that combines the safety principle with Partial Creditability, a view that he calls anti-luck virtue epistemology. In what follows, I will argue that Pritchard's account is not sufficient for knowledge.

8.2.1 *Safety and Partial Creditability Combined*

§ *The insufficiency of safety*. The main reason Pritchard supplements the safety principle with a virtue-theoretic condition is that it cannot rule out cases like Temp as cases of knowledge. Recall the case:

TEMP

> Imagine that our agent—let's call him 'Temp'—forms his beliefs about the temperature in his room by consulting a thermometer on the wall. Unbeknownst to Temp, however, the thermometer is broken and is fluctuating randomly within a given range. Nonetheless, Temp never forms a false belief about the temperature by consulting this thermometer since there is a person hidden in the room, next to the thermostat, whose job it is to ensure that whenever Temp consults the thermometer the temperature in the room corresponds to the reading on the thermometer. (Pritchard *et al.* 2010: 48-49)

Temp's beliefs about the temperature match the truth in the actual and in all close possible worlds, because the hidden agent is so diligent that she would never fail to modify the room's temperature. However, although Temp's beliefs are safe, they do not intuitively amount to knowledge because, unbeknownst to Temp, the world has been changed to make true the content of his beliefs.[5] Pritchard offers the following diagnosis of the case:

> [W]hat is wrong with Temp's beliefs is that they exhibit the *wrong direction of fit* with the facts, for while his beliefs formed on this basis are guaranteed to be true, their correctness has nothing to do with Temp's abilities and everything to do with some feature external to his cognitive agency. (Pritchard 2012b: 260; emphasis mine)

In addition, Pritchard's particular reason for thinking that safety is not sufficient for knowledge is this:

> [The safety principle] will simply demand a match between belief and fact in appropriate counterfactual cases, but a cognitive ability requires far more than this—viz., it requires an appropriate direction of fit between belief and fact. (Pritchard 2012b: 272)[6]

I will accept Pritchard's diagnosis of TEMP as an assumption of my argument against his view. In particular, I will take for granted the following implication:

- DIRECTION OF FIT: S does not know that p if S's belief that p exhibits the wrong direction of fit with the truth-maker of p.

5 Recall again that the hidden agent modifies the temperature of the room and not the values shown on the screen of the thermometer.

6 Sherrilyn Roush (2005: 121-2) makes a related albeit different point: "[t]hough we want safe rather than unsafe beliefs, safety is surely not what makes a true belief knowledge, since knowledge is a matter of responsiveness to the way the world is and safety makes a demand in the opposite direction".

I will also assume that the safety principle is unable to guarantee the direction of fit with the facts that known beliefs have.

§ *The adoption of* PARTIAL CREDITABILITY. Pritchard's way to remedy the insufficiency of safety and to ensure the appropriate direction of fit between belief and fact is to resort to virtue epistemology. In particular, he resorts to PARTIAL CREDITABILITY. Recall the condition:

- PARTIAL CREDITABILITY: *S* knows that *p* only if *S*'s cognitive success is partly or to a significant degree creditable to the exercise of *S*'s cognitive abilities. (Alternatively: only if the exercise of *S*'s cognitive abilities is a salient factor in the total set of factors that explain *S*'s cognitive success).

Pritchard's adoption of PARTIAL CREDITABILITY is motivated by the fact that it is *weak enough* to accommodate cases of knowledge by testimony. Recall that in cases of testimonial knowledge the cognitive abilities of the producer of testimony play an important role in the explanation of the cognitive success of the recipient of testimony.[7] In addition, PARTIAL CREDITABILITY is *strong enough* to rule out cases with the following structure as cases of knowledge:

1. Cases in which *S* forms a belief through a belief-forming process that is considered a cognitive ability but in which the exercise of such a cognitive ability does not explain (not even saliently) why *S* comes to believe the truth (e.g., TEMP).

2. Cases in which the fact that *S* gets things right is not explained by the exercise of cognitive ability because the relevant belief-forming process does not count as a cognitive ability (e.g., cases in which true beliefs arise out of brain lesions or devices capable of generating thoughts that are not yet integrated with the rest of the agent's cognitive dispositions).

The safety principle is still needed to explain why there is no knowledge in some other cases; for example, cases in which one competently and successfully identifies an object (e.g., a barn) but in which one could easily have misidentified a replica of that object (e.g., a fake barn). The resulting account of knowledge combines safety with PARTIAL CREDITABILITY in the following way:

- ANTI-LUCK VIRTUE EPISTEMOLOGY (ALVE): *S* knows that *p* if and only if *S*'s safe true belief that *p* is the product of her relevant cognitive abilities (such that her safe cognitive success is to a significant degree creditable to her cognitive agency). (Pritchard 2012b: 273)

In the next two subsections, I will show that this definition is not sufficient for knowledge.

7 See section 7.4.1.

8.2.2 *The Insufficiency of ALVE (take I)*

The first way of interpreting ALVE is as the conjunction of the safety principle and PARTIAL CREDITABILITY. More specifically, according to this interpretation ALVE says that *S* knows that *p* if and only if:

1. *p* is true,
2. *S* believes that *p*,
3. *S*'s cognitive success (i.e., *S*'s believing truly) is partly or to a significant degree creditable to *S*'s cognitive abilities,
4. *S*'s belief that *p* is safe.

This is the reading of ALVE that Pritchard seems to have in mind: he describes anti-luck virtue epistemology as a view "according to which knowledge is essentially safe true belief *plus* a further epistemic condition (an 'ability' condition), which handles the ability intuition" (Pritchard 2012c: 184, fn. 19; emphasis mine). The ability condition he refers to is PARTIAL CREDITABILITY and the ability intuition that this condition satisfies is the intuition that "knowledge reflects ability, in the sense that when an agent has knowledge, then her cognitive success (i.e., her true belief) is to some significant degree creditable to her cognitive agency (i.e., her exercise of cognitive abilities)" (Pritchard 2012c: 182). Safety, on the other hand, is aimed at satisfying the so-called anti-luck intuition, i.e., the intuition that when one knows, it is not by luck that one comes to believe the truth. Pritchard interprets these two intuitions as imposing two independent constraints on one's theory of knowledge and, consequently, the epistemic conditions aimed at satisfying them need to be independent too, in the sense that there must be at least one type of case that one condition rules out as a case of knowledge that cannot be ruled out by the other condition, and *vice versa*. For example, safety holds in TEMP; PARTIAL CREDITABILITY does not. PARTIAL CREDITABILITY holds in some Gettier-style cases (see below); safety does not.[8] That being so, it is then more congruent with Pritchard's dual approach to knowledge to interpret ALVE as entailing a conjunction of two epistemic conditions than to understand it as entailing a single condition that merges both requirements in a non-conjunctive way.

At any rate, I am not particularly interested in discussing which interpretation of ALVE is the one that Pritchard really has in mind. Pritchard's formulation of ALVE is consistent with two readings (ALVE as the conjunction of safety and PARTIAL CREDITABILITY and ALVE as a non-conjunctive combination of both). Then, my objections against ALVE can be understood in conditional terms: if the former is the

8 See Carter (2013) for an analysis of Pritchard's requirement that the two epistemic conditions aimed at satisfying the two 'master' intuitions must be independent of each other.

correct interpretation of ALVE, then one objection applies; if the latter is the correct interpretation, then another objection applies. As we will see, the conclusion will be the same in either of the two cases: Pritchard's account is not sufficient for knowledge.

§ Kelp's objection to ALVE. Before presenting a couple of counterexamples to the first interpretation of ALVE, let me discuss an objection made by Kelp (2012). According to Kelp, the contribution of Temp's cognitive abilities to his cognitive success parallels the contribution of the cognitive abilities of a person who comes to know the position of a landmark by asking for directions to a stranger, who gives the information correctly. Kelp's point is that if cognitive success is to a significant degree creditable to cognitive ability in the latter case (as Pritchard maintains), so should be in TEMP, and if cognitive success is not to a significant degree creditable to cognitive ability in TEMP, it should not be creditable in the testimony case either. The upshot is, according to Kelp, that while Pritchard can explain each case individually, he cannot explain the two cases in conjunction.

However, there is reason to doubt that the contribution of cognitive ability to cognitive success is the same in both cases. Note that a randomly fluctuating thermometer is not analogous to the reliable informant in the testimony case. To be analogous, the informant should give directions at random, some of them probably making no sense. In view of this disanalogy, it is not clear whether there would be testimonial knowledge if the informant uttered directions at random and the person believed the information.

§ ALVE: *first interpretation, first counterexample*. We can try a better objection to this interpretation of Pritchard's theory. Recall ACTOR, which I used in chapter 7 to discuss the situational view of aptness. Let be *F* the set of features detected by a recipient of testimony that indicate her that her interlocutor is a reliable informant.

> ACTOR
>
> Morris, new in the city, wants to know the position of the Sears Tower. He looks for a good informant and selects Philip, a person who looks like a police officer, on the basis of *F* (*F* may include, for example, those features that make people trust the police in general). However, Philip happens to be an actor cleverly disguised as a police officer whose intention is to give wrong directions to outsiders (*F* indicates reliability but Philip is not reliable). It also happens that whenever Philip is about to lie about the landmark's location, a benevolent Genie moves the Sears Tower so that Philip speaks truly. Consequently Morris's beliefs about the position of the skyscraper are true.

Note that condition (4) (the safety requirement) is satisfied because of the diligent intervention of the Genie. In addition, condition (3) (Partial Creditability) seems to be satisfied too, because there does not seem to be anything in Morris's cognitive performance that is epistemically defective. If Morris performs badly from a cognitive point of view, his cognitive performance is as 'bad' as in the good case where he trusts a real police officer and acquires knowledge. Moreover, Morris selects his informant on a better basis than *S*, an agent who comes to know the location of a landmark by asking the first adult passerby in a train station. While *S* selects an informant on the basis that the person is an adult (rather than a child), Morris selects an informant on the basis that the person is an adult *and* that he looks like a police officer (police officers and taxi drivers are typically the most reliable informants when it comes to giving directions). Nevertheless, intuition says that Morris does not know the location of the skyscraper and consequently ALVE is not sufficient for knowledge.

§ *Possible reply*. It could be replied that Morris *does* know the location of the Sears Tower, provided that he forms his beliefs on grounds that are at least generally appropriate and Morris's interlocutor utters truths. That may well suffice for acquiring knowledge via testimony, one might argue, as knowledge plausibly arises whenever the informational source (e.g., an informant or an epistemic instrument) provides true information and one's ability to detect reliable informational sources is itself (by and large) reliable.

§ *Rejoinder.* However, the problem with this reply is that the direction of fit exhibited by Morris's beliefs seems as wrong as the one exhibited by the target beliefs in Temp. In this way, it should be explained why Actor, in particular, and testimony, in general, are exceptions to Direction of Fit, which as we saw at the beginning is one of Pritchard's core assumptions (and so is for us).

§ ALVE: *first interpretation, second counterexample*. In any case, we can present another counterexample to ALVE in which this kind of reply is ruled out from the outset. The counterexample is built on the following Gettier-style case by Greco (2012):

> Inspired Worker
>
> Jones believes *that someone in the office owns a Ford*, basing his belief on extensive evidence that his co-worker, Nogot, owns a Ford. But Jones's evidence about Nogot is misleading—Nogot owns no Ford. However, another co-worker in the office, Havit, is inspired by Jones's excellent reasoning about Nogot. So much so that he is shaken out of his long

> term depression, and goes out and buys a Ford. (Greco 2012: 14)

According to Greco, the case illustrates why Partial Creditability is not sufficient for knowledge. Since Jones's competent reasoning is the cause Havit buys a Ford and Jones's belief that someone in the office owns a Ford is true because Havit buys a Ford, part of the explanation of why Jones gets things right is the deployment of cognitive ability. Yet Jones does not know that someone in the office owns a Ford.

In a sense, the world changes to make true Jones's belief, i.e., the direction of fit exhibited by Jones's belief seems to be the wrong one. Note that the particularity of the case is that the exercise of cognitive ability causes that the proposition believed has a truth-maker. That is to say, unlike in Actor, the reason the world changes is the exercise of the agent's cognitive abilities and not some external factor (such as the intervention of benevolent Genies). Assuming that this diagnosis is correct, we can transform Inspired Worker into a counterexample to ALVE by tweaking the case in such a way that the safety condition holds (in Inspired Worker, Jones's belief is unsafe):[9]

> Inspired Workers
>
> Jones believes *that someone in the office owns a Ford*, basing his belief on extensive evidence that his co-worker, Nogot, owns a Ford. But Jones's evidence about Nogot is misleading—Nogot owns no Ford. However, unbeknownst to Jones, everyone in the office is inspired by his excellent reasoning about Nogot. So much so that everyone is shaken out of a long term depression, and goes out and buys a car. Curiously, Fords are the only cars that are available for sale on that day.

Jones's actual belief that someone in the office owns a Ford could not easily have been false. After all, if one or two co-workers had refrained from buying a car, the rest of the staff would still have bought one (a Ford). Put it differently, in nearly all (if not all) close possible worlds in which Jones makes the same excellent reasoning about Nogot, it is not the case that everyone in the office does not buy a Ford. Therefore, safety holds. If Greco is right in thinking that Partial Creditability holds as well, then Inspired Workers is a counterexample to the sufficiency of ALVE (see a possible line of response below).

8.2.3 *The Insufficiency of ALVE (take II)*

Another way of interpreting ALVE is as follows. *S* knows that *p* if and only if:

9 See fn. 10 for Pritchard's preferred formulation of safety.

1. *p* is true,
2. *S* believes that *p*,
3. *S*'s safe cognitive success (i.e., *S*'s believing that *p* safely) is partly (or to a significant degree) creditable to *S*'s cognitive abilities (alternatively: *S*'s safe cognitive success is partly (or saliently) explained by the exercise of *S*'s cognitive abilities).

This reading is consistent with the formulation of ALVE quoted earlier. Although this does not seem to be the interpretation of ALVE that Pritchard has in mind, it is worth discussing it because it introduces a novel and powerful idea in the epistemological debate: the requirement that, in order to have propositional knowledge, the *safety* (not just the 'hitting' on truth) of one's beliefs must be *partly* (or *to a significant degree*) creditable to one's cognitive agency. Given Pritchard's own version of the safety principle,[10] this requirement would amount to the condition that, to know that *p*, the exercise of one's cognitive abilities must be a salient factor in the explanation of why the belief that *p* hits the truth in the actual world and in nearly all (if not all) close possible worlds where one forms the belief in the same way as in the actual world.

Condition (3) is not met in ACTOR: the fact that Morris comes to believe the truth in most close possible worlds in which he asks Philip for directions is solely explained by the intervention of the benevolent Genie, who relocates the Sears Tower so that Philip speaks truly. In addition, the requirement is not met in the counterexamples to the sufficiency for knowledge of purely virtue-theoretic accounts (cases of knowledge-undermining environmental luck). Besides, it is not satisfied in counterexamples to accounts that define knowledge just in terms of safety (e.g., TEMP). Finally, the condition holds in cases of inductive knowledge, which prove that the sensitivity principle is not necessary for knowledge.[11] In sum, ALVE, so interpreted, seems to be in a better position than many of its rivals (viz., than 'pure' virtue epistemology and than safety and sensitivity-based theories).

§ ALVE: *second interpretation, first counterexample*. Pritchard could adopt this reading of ALVE to avoid counterexamples like ACTOR, if the first

10 Pritchard formulates several versions of safety. For example, in (2005: 163) he advocates the following version of the principle: "For all agents, φ, if an agent knows a contingent proposition φ, then, in nearly all (if not all) nearby possible worlds in which she forms her belief about φ in the same way as she forms her belief in the actual world, that agent only believes that φ when φ is true". Pritchard judges this version of safety to be superior to a weaker version in which the agent's belief must be true in most close possible worlds. Further reflection on the lottery puzzle leads Pritchard to the adoption of another version in which the target belief must be true in most close possible worlds and in all very close nearby worlds in which the agent forms the belief in the same way as in the actual world (see Pritchard 2012c and section 7.6.2). My objections to ALVE apply independently of the version of safety included in the definition of knowledge.

11 See section 5.3.2.

interpretation is the one that he really has in mind. After all, his formulation of ALVE is consistent with understanding the theory in this way. However, although this reading is better than the conjunctive interpretation, it is not sufficient for knowledge either. To begin with, it seems that condition (3) holds in Inspired Workers. Condition (3) says that known beliefs have the following modal profile: <true in the actual world and true in nearly all (if not all) close possible worlds in which they are formed in the same way as in the actual world>. More importantly, it says that the exercise of the relevant cognitive abilities must be a salient factor in the total set of factors that explain this modal profile. In Inspired Workers, the following holds: in nearly all (if not all) close possible worlds in which Jones believes that someone in the office owns a Ford via the same method of belief formation that he uses in the actual world (reasoning), many workers, inspired by that method, buy Fords, and hence the fact Jones comes to believe the truth in all those worlds is partly because of his cognitive abilities.[12] Of course, Jones does not know that someone in the office owns a Ford.

§ *Possible reply*. In reply, Partial Creditability could be supplemented with some clause restricting the sort of creditability relation that must hold between the exercise of cognitive ability and the fact that the agent gets to the truth about whether p. In Inspired Workers, the fact that Jones gets things right, although creditable to his reasoning abilities, is not *directly* creditable to their exercise, in the sense that it is creditable to the causal impact on the world of the way he forms his belief, and namely to the fact that the exercise of cognitive ability produces a truth-maker of the proposition believed (and not a belief that is true in virtue of some already existing truth-maker). To see this more clearly, think of a method of belief formation as a function that takes as input certain piece of evidence or information and that gives as output a belief that is true or false. The fact that an agent comes to believe the truth is not directly creditable to the operation of that function if the function has a side effect, that is, if in addition to returning a value (the belief), it also modifies the state of the world in such a way that the value is true independently of the evidence or the information taken as input. This line of reply needs to be elaborated, but the general idea should be sufficiently clear: by restricting the creditability requirement of Partial Creditability to something like direct or non-deviant creditability, Partial Creditability (and ALVE) might be able to explain why there is no knowledge in Inspired Worker(s).

§ ALVE: *second interpretation, second counterexample*. As one can imagine, it is not an easy task to design a counterexample to ALVE. Note

12 I talk indistinctively of ways of believing and of methods of belief formation. I will justify this point in chapter 9.

that INSPIRED WORKERS is a case originally constructed to call into question PARTIAL CREDITABILITY that I have subsequently modified so that the safety condition holds. We can proceed the other way around, that is, we can start with a case that challenges the sufficiency of safety (i.e., a case in which the modal profile of the target belief is <true in the actual world and true in nearly all (if not all) close possible worlds in which it is formed in the same way as in the actual world>) and then modify it so that the exercise of cognitive ability meets the demands of PARTIAL CREDITABILITY and of condition (3).

A good candidate is TEMP. Temp's beliefs are safe, but their safety is completely explained by the manipulation of the hidden agent, so that we cannot credit him with knowledge. For example, if the temperature in the room is 25°C and the broken thermometer reads 25.4°C, whenever Temp is about to form the belief that the temperature is 25.4°C by consulting the thermometer, the hidden agent ensures that temperature of the room reaches 25.4°C so that it matches the value shown on the screen. This manipulation takes place in the actual world and in all close possible worlds and, therefore, safety holds.

What we seek next is a case in which the exercise of cognitive ability is a salient factor in the explanation of the safe modal profile of the target belief, while we can still retain the intuition that knowledge is absent because an aspect of the situation, which is beyond the agent's cognitive reach, partially explains, together with her relevant cognitive abilities, the safety of her cognitive success. As in TEMP, this factor will be the manipulation of a hidden person, but this time the manipulated agent (and the case) will be called Tempy:

> TEMPY
>
> THE AGENT: Tempy was born with an extraordinary although slightly inaccurate ability to detect the temperature: she can perceive that the temperature of the environment is in a range of ± 1°C in a very reliable manner. For example, if the temperature is 25.4°C, she can tell that it is between 24°C and 26°C, but she cannot know that the temperature is exactly 25.4°C.
>
> THE LOCATION: She is in her room, where she does not normally use her detection ability because it requires a lot of effort and concentration. Instead, she usually looks at the digital thermometer on the wall to know the temperature of the room (as in TEMP, the thermometer has always been very reliable in the past and by consulting it the agent has acquired knowledge many times).
>
> THE EPISTEMIC POSITION: This time, Tempy has acquired, via a trustworthy informant (e.g., the heating technician), excellent justification to believe the following *false* propo-

sition: <the thermometer indicates 3°C more than the real temperature>. That is, unless Tempy exercises her detection ability, she will believe that the temperature of the room is 3°C more than what the thermometer says.

The special circumstances: Unbeknownst to Tempy, the thermostat is broken and fluctuating randomly within a range of ± 0.5ºC with respect to the real temperature and, as in Temp, a hidden agent raises or lowers the temperature of the room so as to match the values shown on the screen. This agent can raise or lower the temperature of the room only within the range of fluctuation of the thermometer (± 0.5ºC).

The belief-forming method: Since Tempy has reasons to distrust the thermometer, she decides to implement the following belief-forming method: first she will use her detection ability to check in which range the real temperature is; then, she will look at the thermometer to check whether it is true that it indicates 3°C more than the real temperature (as she has been told). If this is false, then, given that the thermometer has always been very reliable in the past, she will finally trust the readings of the thermometer.

The formation of the belief: Tempy perceives that the temperature is between 24ºC and 26ºC (the real temperature is 25.4ºC). The broken thermometer says that it is 25ºC. As a consequence, Tempy rejects the justification to believe that the temperature is 28ºC. Then, by trusting the thermometer's reading, she forms the belief that the temperature is 25ºC, but here is the trick: in the meantime, unbeknownst to Tempy, the hidden agent has adjusted the temperature from 25.4ºC to 25ºC.

The intuition: Tempy's belief that the temperature is 25ºC is true, but it does not amount to knowledge.

The case is complex but the explanation of why it is a counterexample to ALVE is simple. Without the detection ability, Tempy (or any other person) would have believed that the temperature is 28ºC. Why? Because of the excellent testimonial justification to believe the *false* proposition that the thermometer indicates 3°C more than the real temperature (it actually indicates 25ºC). This means that without the detection ability in place the target belief would have had the following modal profile: <false in the actual world and false in all close possible worlds in which it is formed in the same way as in the actual world>.[13] However, given that the use of the detection

13 Note that the case is described in such a way that the hidden agent would have adjusted the temperature of the room to 25ºC (the reading of the thermometer) re-

ability is part of Tempy's belief-forming method, she is in a position to reject her testimonial justification and, consequently, her belief is true in the actual world, but also, and more importantly, in all close possible worlds. That is, Tempy's cognitive abilities *saliently* (but do not completely) explain the safety of her cognitive success. Therefore, condition (3) holds (Partial Creditability obviously holds as well).

Note that the same diagnosis offered by Pritchard of why Temp is not a case of knowledge applies here. As in Temp, the direction of fit exhibited by Tempy's belief is the wrong one: although there is correspondence between belief and fact, the world is changed by the hidden agent to match the content of the belief. Following the dictate of Direction of Fit, Tempy does not know. Therefore, Pritchard's diagnosis of Temp, fully applicable to Tempy, indicates that ALVE cannot guarantee that known beliefs have an appropriate direction of fit with the facts. Ironically, the preservation of the appropriate direction of fit was one of Pritchard's main motivations for the adoption of a virtue-theoretic condition in the first place. However, the weak virtue-theoretic condition that Pritchard adopts (Partial Creditability) is *too* weak.

One might still have the intuition that Tempy knows that the temperature is 25°C degrees, but this thought arises, I think, because the difference between the range of temperature that Tempy can detect (± 1°C) and the range of fluctuation of the thermometer (± 0.5°C) is not very significant. Accordingly, one may feel that, given the degree of accuracy of Tempy's detection ability, her belief that the temperature is 25°C has an appropriate direction of fit and that, although the world is certainly changed to match the content of the belief (from 25.4°C to 25°C), such a change is not significant enough to reverse its direction of fit with the facts.

We can strengthen the intuition that the direction of fit is the wrong one by increasing the difference between the range of temperature that Tempy can detect and the range of fluctuation of the thermometer. For example, suppose that: 1) Tempy is able to perceive that the temperature of the environment is in a range of ± 7°C in a very reliable manner; 2) she has justification to believe that the thermometer indicates 10°C more than the real temperature; 3) the range of fluctuation of the thermostat is ± 2°C with respect to the real temperature; 4) the temperature in the room is 30°C; 5) the broken thermometer says that it is 28.5°C; 6) unbeknownst to Tempy, the hidden agent lowers the temperature from 30°C to 28.5°C. In this case, Tempy perceives (knows) that the temperature is between 37°C and 23°C (the real temperature is 30°C). Since the broken thermometer says that it is 28.5°C, she rejects the justification to believe that the temperature is 38.5°C and comes to believe (falsely) that the thermometer is working prop-

gardless of Tempy's evidence or of the reliability or accuracy of her belief-forming method.

erly. Finally, when she is about to form the belief that the temperature is 28.5ºC, the hidden agent adjusts the temperature from 30ºC to 28.5º C. Clearly, Tempy does not know that the temperature is 28.5ºC. In what follows, I will evaluate two possible replies.

§ *Possible reply (I)*. It could be argued that Tempy's belief-forming process belongs to the same category as the belief-forming processes of Laurence BonJour's clairvoyants (BonJour 1985: 38-41) and Keith Lehrer's Mr. Truetemp (Lehrer 1990: 163-164).[14] Mr. Truetemp, for instance, is an agent with a small device implanted in his brain that can detect with great precision the temperature of the environment, and that is also capable of generating beliefs about the temperature on that basis. Lehrer thinks that although the device is a very reliable way of forming true beliefs about the temperature, Mr. Truetemp's beliefs do not qualify as knowledge. Many epistemologists share this intuition and, consequently, many epistemologists will have the same intuition about Tempy *if* Tempy's belief-forming process is like Mr. Truetemp's. Pritchard could avail himself of that intuition and argue that Tempy's safe cognitive success is not creditable to cognitive ability because there would not be cognitive ability in the first place.

§ *Rejoinder.* As Lehrer describes him, Mr. Truetemp has no idea that the device has been inserted in his brain and that he is able to detect the temperature. In fact, he is puzzled about why he thinks so obsessively about the temperature. As we saw in chapter 7, this can be explained in terms of a lack of cognitive integration: Mr. Truetemp's disposition to form beliefs in that way is simply not integrated with the rest of his cognitive dispositions. Recall the type of conditions that are thought necessary for cognitive integration. *S*'s cognitive disposition is cognitively integrated with other of *S*'s cognitive dispositions only if: 1) it is sufficiently stable; 2) it causally interacts with the other dispositions; 3) the beliefs it produces cohere with the beliefs produced by the other dispositions; 4) the whole system of *S*'s cognitive dispositions is disposed to form and maintain a coherent set of beliefs; 5) *S* has taken responsibility for the cognitive disposition; 6) *S* has a reflective perspective over her disposition.

As I explained, there is some disagreement on whether some of these requirements are necessary for cognitive integration or not, but this does not matter for our purposes here, as we can modify the details of Tempy in such a way that Tempy's detection ability satisfies whichever conditions are considered necessary for cognitive integration: from the more internalist requirements (coherence, reflective perspective on one's cognitive disposition, awareness of its reliability, and so on) to the more externalist demands (stability and causal interaction requirements, implicit recognition that one is the source of one's beliefs, and so on). By way of illustration, we can stipulate that

14 See section 7.3.

Tempy's cognitive ability is as cognitively integrated as Tempo's.[15] Recall how Pritchard describes him:

> [Tempo] is fitted from birth with a highly reliable device which records the ambient temperature and Tempo grows up in a culture where it is taken for granted that one consults one's temperature-recording device in order to form beliefs about the ambient temperature. (Pritchard 2010b: 146)

Pritchard asks: "Wouldn't we nonetheless straightforwardly regard him as gaining knowledge via this belief-forming process? Moreover, wouldn't we regard Tempo's cognitive success as being to a significant degree creditable to his cognitive agency (...)?" (*ibid*.). The same rhetorical question can be asked about Tempy.

§ *Possible reply (II)*. The second possible reply is the following: it could be argued that Tempy's abilities saliently explain that she truly believes that the temperature is between 24ºC and 26ºC and consequently that she does not falsely believe that it is 28ºC. However, they do not saliently explain that she truly believes that the temperature is 25ºC. In other words, they do not saliently explain that she attains cognitive success (and *a fortiori* that she attains *safe* cognitive success).

§ *Rejoinder.* The problem with this reply is that it only works on the assumption that the fact that Tempy manages not to believe falsely that the temperature is 28ºC does *not* contribute to the explanation of why she comes to believe truly that the temperature is 25ºC. But note that, as the case is described, Tempy's excellent testimonial justification is a defeater for the belief that the temperature is 25ºC. Tempy's beliefs are based on the readings of the thermometer. Therefore, unless Tempy rejects her justification to believe that the thermometer indicates 3ºC more than the real temperature, she will never come to believe the truth. In this sense, Tempy's cognitive abilities and namely the belief-forming method that she implements saliently explain why she attains *actual* cognitive success. Of course, her cognitive success is also saliently explained by the intervention of the hidden agent. But since partial contribution to cognitive success is all we need, the case is a counterexample to ALVE (to both interpretations).

Nevertheless, for those who are not fully convinced yet of the fact that if Tempy had not rejected her good justification to believe that the thermometer indicates 3ºC than the real temperature, she would not have believed the truth, we can strengthen the intuition that the rejection of such a good justification saliently (or partly) contributes to her actual cognitive success by slightly modifying the case. Let us stipulate that the hidden agent would intervene only if Tempy

15 See TEMPO in section 7.3.

exercised her detection ability. Thus, the fact that Tempy exercises her detection ability explains not only the rejection of her justified false belief that the thermometer indicates 3ºC more than the real temperature, but also the intervention of the hidden agent. In this way, Tempy's cognitive abilities are a salient factor in the total set of factors that explain her cognitive success. The other salient factor is, of course, the manipulation of the hidden agent, which is precisely what prevents Tempy's belief from being knowledge.

8.3 THE DIRECTION OF FIT PROBLEM

For the sake of completeness, I will offer some considerations on what is exactly what TEMPY does and does not show and on the reach of the challenge that the case poses. This will also serve to expand my account of the situational view of APTNESS. First, the case shows that ALVE is insufficient for knowledge, but it does not show that the whole approach of combining modal conditions (e.g., safety) with virtue-theoretic requirements is misguided. There are other accounts that can (arguably) explain the case (e.g., Kelp 2012; Turri 2011; *forthcoming*). Second, TEMPY does not settle the question neither in favor of an interpretation of the 'because of' relation in terms of manifestation of competence (Kelp 2012; Sosa 2011; Turri 2011), nor in favor of an interpretation in terms of creditability/explanatory salience (Greco 2010; Pritchard 2012b; Sosa 2007). Nevertheless, strong virtue-theoretic conditions in terms of creditability/explanatory salience (Greco's FULL CREDITABILITY) seem to be in a better position to explain TEMPY than strong virtue-theoretic conditions in terms of manifestation of competence (Sosa's aptness, Turri's adeptness). We will see in a moment how these strong virtue epistemologies explain the case. First, let me state the challenge that cases like TEMPY present.

§ *The direction of fit problem.* Pritchard is right in thinking that what is epistemically problematic in TEMP is that Temp's beliefs exhibit the wrong direction of fit with the facts. As I argued, Pritchard's diagnosis of TEMP can be extrapolated to TEMPY, ACTOR and, plausibly, INSPIRED WORKER(S). Therefore, the challenge posed by cases like TEMP, TEMPY, ACTOR and INSPIRED WORKER(S) is that of specifying what is required to preserve the appropriate direction of fit that known beliefs have. Importantly, note that not only known beliefs exhibit an appropriate direction of fit (to have an appropriate direction of fit just seems a necessary condition for knowledge). Plausibly, in certain cases of knowledge-undermining luck the direction of match between facts and luckily true beliefs is the one that befits knowledge. Consider the following case:

PICASSO

> There is a temporary exhibition of Picasso's Blue Period at the local gallery. George, a Picasso enthusiast, thinks that all the works of that period are great and, as soon as he reads about the exhibition, he goes to the gallery. He enters the first exhibition room and starts looking at the paintings that are on his right. He normally proceeds clockwise but for some improbable reason he has decided to start counterclockwise. He looks at the first painting and thinks "The painting I am looking at is a great Picasso painting". Unbeknownst to George, the director of the exhibition has decided to play a joke on a pretentious local art critic who boasts about his vast knowledge of Picasso's work replacing most original Picasso paintings with masterful Picasso fakes that only an expert eye could detect. George is a Picasso enthusiast, but not an expert, and he has looked at the only original Picasso painting in the room.[16]

Many epistemologists would claim that George does not know that the painting he is looking at is a great *Picasso* painting because in most close possible worlds in which he believes the same proposition, that proposition is false. Note that George could easily have started looking at the paintings on his left, and given that he thinks that all the works of the Blue Period are great, he could easily have believed falsely that the paintings on his left are great Picasso paintings. Therefore, it is by luck that George's actual belief is true. As we have seen in previous chapters, the kind of veritic luck instantiated is environmental luck (features of the environment would make the target belief fail to hit the truth in close possible worlds *but not in the actual world*).

Crucially, the direction of fit of George's belief seems to be the appropriate one, in the sense that it is as appropriate as the direction of fit of an ordinary known perceptual belief (possibly because in both cases there are no non-deviant causal connections between beliefs and facts). In view of this, the *direction of fit problem* is, then, the problem of how to spell out an epistemic condition that holds (1) in cases of knowledge *and* (2) in cases of environmental luck but that does not hold (3) in cases like TEMP, TEMPY, ACTOR and INSPIRED WORKER(S).

8.3.1 *Full Creditability*

Virtue-theoretic conditions, in general, and strong virtue-theoretic conditions (FULL CREDITABILITY, APTNESS) in particular, seem to be the sort of epistemic requirements best suited to solve the direction of fit

16 PICASSO is structurally equivalent to Ginet/Goldman's FAKE BARNS (Goldman 1976).

problem. Let us start by analyzing whether Greco's Full Creditability is able to solve it. Recall it:

- Full Creditability: *S* knows that *p* if and only if *S*'s cognitive success is *primarily creditable* to *S*'s cognitive abilities (alternatively: if and only if the exercise of *S*'s cognitive abilities is the *most salient* factor in the total set of factors that explain *S*'s cognitive success).

Full Creditability is satisfied in cases in which an agent knows by means of reliable cognitive faculties such as perception, reasoning or memory. Arguably, it is also satisfied in cases of environmental luck like Picasso (this is one of the two objections that Pritchard makes against Greco's account of knowledge). In addition, it does not hold in cases like Temp or Tempy, because Temp and Tempy's cognitive successes are (respectively) non-creditable and just partly creditable to their cognitive abilities. Full Creditability is not satisfied in Actor or in Inspired Worker(s) for similar reasons. Therefore, it seems that it is suitable to solve the direction of fit problem. However, we saw in chapter 7 that Full Creditability rules out cases of testimonial knowledge as cases of knowledge (this is the second objection that Pritchard makes against Greco). Therefore, Greco's account fails to solve the direction of fit problem.

8.3.2 *Aptness (Appropriateness View)*

Let us evaluate whether virtue-theoretic conditions formulated in terms of manifestation of competence can solve the problem. Recall Aptness, which is defended by Kelp (2012), Sosa (2007, 2011) and Turri (2011):

- Aptness: If *S* knows that *p*, *S*'s cognitive success (i.e., *S*'s believing truly that *p*) *manifests* epistemic competence (or cognitive ability).

As explained in chapter 7, the way these commentators think about knowledge is analogous to the way we think about dispositions. In the same way as salt manifests its solubility when stirred into water, a belief manifests cognitive competence when it hits the truth. Now, salt can only manifest its solubility under some 'normal' or 'appropriate' conditions, and namely when it is introduced in a liquid with such-and-such characteristics. In the same way, beliefs can only manifest cognitive competence when they hit the truth under some conditions. In particular, the agent's internal conditions and the external conditions of the environment or the situation must be intuitively appropriate. Only under such conditions, they argue, beliefs can be assessed for aptness. I called this the *appropriateness view of aptness*.

For example, when we look at objects under good light conditions we can form apt beliefs about those objects. Arguably, we can also form apt beliefs via testimony when our interlocutors are globally reliable (in Actor, for example, the interlocutor is an expert liar). In appropriate circumstances for testimony, our believing the relevant truths manifests our competence for selecting reliable informants (if we have it in the first place, of course). In addition, in cases of environmental luck the target beliefs are also apt because the conditions are considered appropriate and competence is manifested under such conditions. For example, in Picasso, George looks under good light conditions at a real Picasso painting (not to a fake or to a hologram), which means that the circumstances are as appropriate as the circumstances in which he comes to know some other proposition by visually perceiving an object. In sum, Aptness (interpreted according to the appropriateness view) seems a suitable condition to solve the direction of fit problem: it holds in cases of knowledge (testimonial knowledge included) and in cases of environmental luck

Is Aptness (interpreted in that way) suitable to explain why the target beliefs in cases like Temp or Tempy exhibit the wrong direction of fit? As in Actor, the way in which advocates of this view of Aptness would explain why Temp and Tempy are not cases of knowledge would be to point out that the requirement for apt belief that the agent must be appropriately situated is not met. They would argue, for instance, that the unreliability of the thermometer makes the situation inappropriate, so that Temp and Tempy's cognitive successes do not manifest competence in those specific environments.[17]

However, we saw in chapter 7 that this conclusion is *ad hoc* in a modified version of Temp in which the thermometer randomly fluctuates and subsequently works properly at consecutive short intervals. Cases of this sort undermine the strategy of the advocates of the appropriateness view of splitting the circumstances into two categories: appropriate and inappropriate. Consequently, I do not think that they have a satisfactory explanation of why the direction of fit of the target beliefs of these cases is the wrong one. This is even more evident once we compare their treatment of the cases with Greco's clear and straightforward explanation that the right direction of fit between belief and fact can only be preserved if the agent's cognitive success is primarily creditable to her cognitive abilities (an explanation that is unsatisfactory for other reasons, as we have seen).[18]

---

17 In the case of Tempy, it seems that her cognitive success does manifest her ability to detect the temperature (at least partially). This fact puts even more pressure on advocates of the appropriateness view of Aptness.

18 Interestingly, the sensitivity principle explains why Temp and Tempy's beliefs exhibit the wrong direction of fit: they are unresponsive to the world, i.e., insensitive. In particular, they are insensitive because both Temp and Tempy would continue to believe the target propositions in the closest possible worlds where the target beliefs are false (worlds where the hidden agents do not intervene). However, as Sosa (1999) and Vogel (2007) show, beliefs that are known by induction are not sensitive. Therefore, either sensitivity does not solve the direction of fit problem or inductive

8.3.3 *Aptness (Situational View)*

The situational view of aptness gives a satisfactory answer to the direction of fit problem. According to this view, the degree of competence that an agent must have in order for her success to manifest competence (i.e., in order for her successful performance to be apt) varies as a function of the environment. Recall the explanation given in chapter 7: for each situation *S* there is a minimum level *L* of competence *C* that an agent possessing *C* must reach in order for her successful performance to manifest *C* (*L* is determined by pragmatic factors and becomes evident when we compare cases of the same type). There is no split between appropriate or normal circumstances and inappropriate or abnormal ones: all that matters is that one has the relevant competence and that one is able to exercise it with certain level of proficiency.

In chapter 7, I explained how the situational view rules out Actor, Temp and the short 'interval' version of Temp as cases of knowledge. Roughly, in these cases the level of competence required for manifesting competence is very high and the agents in question, although competent, do not reach it. In the 'long interval' version of Temp, where Temp comes to know the temperature by consulting the thermometer, the level of competence required for aptness is the usual one, namely Temp only needs to trust the device competently.

In Tempy, the level of competence required for manifesting competence is even higher than in Temp: in addition to trusting the thermometer properly and detecting the abnormal changes of temperature in the environment, Tempy needs to reject the good justification she has to believe the false proposition that thermometer indicates 3ºC more than the real temperature. While Tempy is competent enough to reach the level of competence required for the latter task (thanks to her detection ability), she is not able to reach the level of competence required for the former: to detect the the abnormal changes of temperature she would need, for example, a very fined-tuned ability to detect the temperature. For this reason, her belief that the temperature is 25ºC, although quite competent, is not apt *in that situation* (in other situation it could be apt with the same level of competence displayed).

8.3.3.1 *Inferential Belief Formation*

Is the situational view of aptness able to explain Inspired Workers and thus give a solution to the direction of fit problem? Inspired Workers, in particular, and cases of inferential belief formation, in general, pose a challenge to aptness-based accounts. Greco (2012) states the challenge in the following way:

knowledge is an exception to the direction of fit requirement. I leave the question open.

> [I]t is unclear how or why [an aptness-based account] distinguishes between reasoning based on misleading evidence, as in some standard Gettier cases, and reasoning based on knowledge-producing evidence. For example, consider the case where Jones reasons on the basis of non-misleading evidence that someone in his office owns a Ford. Does true belief here manifest Jones's reasoning competence? Let's say it does. But then why does it not in the Gettier case, where Jones reasons just as well, and by doing so ends up with a true belief? Remember, Jones makes no mistake in reasoning in the Gettier case—each step of his reasoning, including the final existential generalization, is flawless. (Greco 2012: 8)

Obviously, in order for one's inferential true belief to manifest one's reasoning competence (and hence to qualify for knowledge) one needs to infer the belief competently. But I think that something like the following condition must hold as well:

- *S* knows that *q* by competently deducing *q* from *p*, only if *S*'s acquisition of her evidence for *p* manifests competence (i.e., only if it is apt).

When does *S*'s acquisition of certain piece of evidence fail to manifest competence? For instance, when *S* has the disposition to acquire evidence from untrustworthy sources and acquires evidence from such a type of source (e.g., when *S* accepts a blatant lie from a bald-faced liar as evidence for *p*). But in some other situations *S* might have the disposition to acquire evidence only from trustworthy sources and yet the circumstances might be such that what looks like a trustworthy source provides in reality misleading evidence (e.g., *S* accepts a report from a usually reliable newspaper that this time has included misleading information with the only purpose of deceiving her). In those cases, the level of competence that is required for *S*'s acquisition of proper evidence to manifest her competence to acquire proper evidence is higher than the maximum level of competence that *S* is able to display (the situational view of aptness also applies to apt acquisition of evidence).

In Inspired Workers, we are only told that Jones bases his belief that someone in the office owns a Ford on extensive evidence that his co-worker, Nogot, owns a Ford. Since the evidence is extensive but misleading, we may suppose that Jones's circumstances are such that is extremely difficult for Jones to find out whether Nogot owns a Ford. In particular, the level of competence required in order for Jones's acquisition of evidence to manifest competence is very high *in that particular situation*. Since Jones does not acquire his evidence for the premise belief in an apt manner, he does not know the conclusion of his deduction that someone in the office owns a Ford.

Note that I do not claim that in order for *S* to know that *q* by competent deduction from *p* to *q* *S* must necessarily acquire true evidence for *p*: *S*'s acquisition of evidence for *p* might manifest competence even though *S*'s evidence for *p* is false. Consider the following case by Warfield (2005):

> HANDOUTS
>
> Counting with some care the number of people present at my talk, I reason: 'There are 53 people at my talk; therefore my 100 handout copies are sufficient'. My premise is false. There are 52 people in attendance—I double counted one person who changed seats during the count. And yet I know my conclusion. (Warfield 2005: 407-408)

How can it be that I manifest competence in acquiring evidence for my premise belief if I have double counted one person and as a consequence my evidence is false? Although perhaps I am not competent in counting people, I may have the ability to gauge a margin of error for my counting up to a number around 50 (a number significantly smaller than 100, the number of handouts).[19] In this way, I manifest competence in acquiring the relevant evidence even though my evidence is, as a matter of fact, false.

Similar considerations apply to other cases of knowledge by inference from falsehood. Consider what Turri (2011) says in this passage:

> You can proceed competently despite relying on false premises. Falsehood in the form of idealization pervades scientific theorizing and reasoning, much of which is competent and confers knowledge. (...) And for some purposes it doesn't matter if we believe that the gravitational constant is exactly, as opposed to approximately, 6.7 x 10^{-11} $m^3kg^{-1}s^{-2}$ or that π equals exactly 3.14. We might nevertheless reason from these false premises to reach a true conclusion, which outcome would manifest competence. For instance, by relying on that value for the gravitational constant, we could come to know that within the next thousand years the Moon will not crash into Earth due to Earth's gravity. Or by relying on that value for π, we could come to know that a ten-meter-diameter circle has an area greater than fifty square meters. (Turri 2011: 8)

To what Turri says here we should add that, to know in the way described, the idealization itself must manifest competence, i.e., it must be apt. To illustrate: 3.14 is an apt idealization of π; by contrast, idealizing π as 3.15 does not manifest competence (at least in standard contexts such as rounding off in math class). In the same way, 6.7 x

19 I owe this point to Chris Kelp.

10^{-11} $m^3kg^{-1}s^{-2}$ is an apt idealization of *G*; by contrast, idealizing *G* as 6.6 x 10^{-11} $m^3kg^{-1}s^{-2}$ does not manifest competence.

The rough account of inferential knowledge in terms of manifestation of competence just sketched is not intended to cover all instances of inferential knowledge. Rather, it is intended to give a general idea of how aptness-based accounts (and in particular the situational account) can accommodate cases like INSPIRED WORKER(s) or Gettier's original cases. Since APTNESS, interpreted according to the situational view, is able to deliver the correct result in all the cases discussed thus far (cases of non-inferential knowledge, ACTOR, TEMP, the interval version of TEMP, TEMPY, INSPIRED WORKERS, cases of knowledge by inference from falsehood, and so on), we can conclude that it solves the direction of fit problem. That is, APTNESS, so interpreted, guarantees the direction of fit that known beliefs have.

8.4 AMPLITUDE

Turri (2011; *forthcoming*) extends Sosa's AAA performance-assessment structure with another normative property: *amplitude*. According to Turri, a performance is *ample* just in case its safety (not just its accuracy) manifests the agent's competence. He calls the outcome of an ample performance an ample achievement and he takes ample performances and ample achievements to be a subset of apt performances and apt achievements. Turri's account of knowledge is a natural application of this normative framework. He defines ample belief in the following way:

- AMPLE BELIEF: *S*'s belief is ample if and only if its safety (not just its 'hitting' the truth) manifests *S*'s epistemic competences.

His view on knowledge, which he calls the *ample achievement account of knowledge*, or AA+ for short, is modeled on that definition of ample belief:

- AA+: *S*'s knows that *p* if and only if *S*'s belief that *p* is ample.

AA+ has many virtues. It can rule out as cases of knowledge all the cases that safety and aptness are able to explain separately, but more importantly, it rules out the modified versions of FAKE BARNS (FAKE BARNS II and III) as cases of knowledge. In both cases, the safety of Henry's belief that some object in the field is a barn does not seem to manifest epistemic competence (the same explanation applies to TEMPY).

8.4.1 *Two Concerns About Amplitude*

Although I am quite sympathetic to AA+, I have a couple of concerns.

8.4.1.1 *First Concern*

Although AA+ is an intuitive account of knowledge, I think it leaves too much to intuition when it comes to explaining the cases. For instance, cases like TEMP seem paradigmatic examples of true belief whose safety does not manifest competence. However, how does it explain cases like DACHSHUND? Oscar comes to know, in the actual world, that the object over there is a dog. Therefore, his belief is allegedly ample according to AA+. If Oscar could easily have believed falsely that the object over there is a dog, why is Oscar's belief safe? In which sense Oscar's safety would manifest competence given such an easy possibility? The answer is not clear. We need an account of knowledge that gives clear answers to these questions.

8.4.1.2 *Second Concern*

My second concern has to do with Turri's motivation for AA+. As we have seen, Turri thinks that ample achievements are a subset of apt achievements. Turri claims that when one wins a competition and one's victory manifests skill, that victory is an achievement (an apt one). However, one's victory would not count as an ample achievement unless, in addition, one wins in an overwhelming manner, i.e., unless one wins by a wide margin (he gives the examples of Usain Bolt or Michael Phelps' victories at the 2008 Olympic Games in comparison with, say, the victories of a regular runner and a regular swimmer).[20]

Some cognitive achievements certainly seem overwhelming, as when a plain spotter is able to identify a German plane from a great distance, or when someone is able to do mentally a complex multiplication in a couple of seconds. But some other cognitive achievements do not seem overwhelming at all. In DACHSHUND, Oscar could easily have mistaken a wolf for a dog, and yet he knows that the object over there is a dog. Or consider the child who comes to know for the first time that 2 +2 = 4. Such cognitive successes do not seem overwhelming in any sense. So, why should knowledge be considered an overwhelming cognitive achievement? The motivation of AA+ is not accurate.

8.5 SUMMARY

In this chapter, I have explored several ways in which safety might be combined with virtue-theoretic conditions such as aptness or PARTIAL CREDITABILITY in a theory of knowledge. The underlying idea has

20 The apt victory of a runner who wins a race is safe if, unbeknownst to her, the rest of the runners have been drugged for her benefit, but it is not ample (like Usain Bolt's) because the safety of the victory does not manifests the runner's athletic skills, but the skill of whoever has drugged the rest of the runners in her favor.

been that epistemic control does not merely require the conjunction of safety and aptness, but a proper non-conjunctive combination of both. My motivation for this idea has been the intuition that in most cases control over performance arises when one's success is safe *because of* one's abilities.

In section 8.1, I have analyzed the view that knowledge is safe, apt belief (Christoph Kelp's safe-apt view). I have argued that Kelp's way of individuating methods is problematic, but more importantly, that a belief might be safe, apt and yet not knowledge. In section 8.2, I have dissected Duncan Pritchard's anti-luck virtue epistemology (ALVE). ALVE is the view that knowledge is safe true belief partially creditable to cognitive ability. ALVE, I have argued, can be interpreted in two ways: either as the conjunction of safety and PARTIAL CREDITABILITY or as the condition that the safety of one's beliefs must be creditable to cognitive ability.

In section 8.3, I have paused my discussion on how to combine safety with virtue-theoretic conditions to introduce what I call the direction of fit problem in epistemology, the problem of explaining why in cases of knowledge and in cases of environmental luck the direction of fit of the beliefs in question is the appropriate one while explaining why in cases like TEMP it is not. The conclusion has been that only the situational view of aptness can solve the problem.

Finally, in section 8.4 I have examined John Turri's proposal of knowledge as ample belief, i.e., belief whose safety manifests competence. Against Turri's view, I have simply pointed out that it lacks specificity and that it offers a wrong picture about knowledge (as an overwhelming cognitive achievement). In the next chapter, I will propose an account of knowledge closely related to Turri's but that is much more specific on how the safety condition applies to the cases and on the relation between safety and aptness. In addition, its motivation is that of the control theory of knowledge, i.e., it conceives knowledge as a complete cognitive achievement where the notion of complete cognitive achievement is cashed out in terms of epistemic control. In this sense, knowledge is not overwhelming success, just a controlled one.

9

SECURITY: APTNESS IN A SAFE BUBBLE

9.1 A DEFENSE OF THE SAFETY PRINCIPLE

In chapter 5, we saw a series of cases that several commentators (Comesaña 2005; Kelp 2009; Neta & Rohrbaugh 2004) have offered with the aim of showing that unsafe knowledge is possible (i.e., that safety is not necessary for knowledge). Following Tomas Bogardus (2012), we saw that a flaw in the argumentation provided by these authors is that from the fact that *S* was at epistemic risk just before forming her belief they infer the conclusion that *S*'s belief is formed unsafely. In other words, the cases are not counterexamples to safety because they locate epistemic risk *before* the target beliefs are formed.

According to Bogardus, a genuine counterexample to the necessity of safety would be a case in which the agent in question is at epistemic risk *when* she forms the relevant belief and not merely before. And he proposes the following case:

> Atomic Clock
>
> [T]he world's most accurate clock hangs in Smith's office at a cereal factory, and Smith knows this. The clock's accuracy is due to a clever radiation sensor, which keeps time by detecting the transition between two energy levels in cesium-133 atoms. This radiation sensor is very sensitive, however, and could easily malfunction if a radioactive isotope were to decay in the vicinity (a very unlikely event, given that Smith works in a cereal factory).
>
> This morning, against the odds, someone did in fact leave a small amount of a radioactive isotope near the world's most accurate clock in Smith's office. This alien isotope has a relatively short half-life, but—quite improbably—it has not yet decayed at all. It is 8:20 am. The alien isotope will decay at any moment, but it is indeterminate when exactly it will decay. Whenever it does, it will disrupt the clock's sensor, and freeze the clock on the reading "8:22." (Don't ask why; it's complicated.) Therefore, though it is currently functioning properly, the clock's sensor is not safe. The clock is in danger of stopping at any moment, even while it currently continues to be the world's most accurate clock.
>
> Smith is quite punctual, and virtually always arrives in her office on workdays between 8:20 and 8:25 am, though

> no particular time in that duration is more likely than any other to see her arrive. Upon entering her office, Smith always looks up at her clock and notes the time of her arrival. Today, in the actual world (''@''), that alien isotope has not yet decayed, and so the clock is running normally at 8:22 am when Smith enters her office. Smith takes a good hard look at the world's most accurate clock—what she knows is an extremely well-designed clock that has never been tampered with—and forms the true belief that it is 8:22 am. (Bogardus 2012: 12-13)

In this section, I will show two things. In particular, that Atomic Clock does not prove that knowledge and safety can come apart, as Bogardus thinks; in general, that known beliefs are safe even if the agents in question are at high epistemic risk when they form them.[1]

9.1.1 *Setting the Scene*

For the sake of the argument, I will use Duncan Pritchard's version of the safety condition (Pritchard 2005: 163). The following definition of safe belief is based on that version:

- Safe Belief: *S*'s belief that *p* formed in the actual world @ via a belief-forming method of type *M* is safe if and only if (i) it is true in @ and (ii) in nearly all, if not all, close possible worlds in which *S* forms the belief that *p* via a belief-forming method of type *M*, that belief is true.[2]

The safety principle says that if *S* knows that *p*, then *S*'s belief that *p* is safe. Bogardus's argument against the necessity of safety is as simple as follows:

(1) Smith knows that it is 8:22 am via certain belief-forming method (e.g., looking at the clock) in @.

(2) Smith believes falsely that it is 8:22 am via the same type of belief-forming method in very many close possible worlds (i.e., Smith's belief is unsafe).

therefore, unsafe knowledge is possible.

(1) is not only intuitively true, but it is also considered true by many respectable theories of knowledge.[3] Therefore, we must grant that

1 Integral parts of section 9.1 appear in Broncano-Berrocal (*forthcoming*).
2 This definition only takes into account one sense of knowledge-undermining risk. A complete definition of safe belief should also take into account the risk of believing a closely related proposition, a proposition with gappy content, with paradoxical content or with no content at all. For the sake of simplicity, I will use this incomplete conception of safe belief to discuss Bogardus's criticism.
3 Bogardus is particularly interested in showing the truth of (1) and reviews, accordingly, a representative set of theories.

Smith knows that it is 8:22 am when she looks at the clock in @. Let us concentrate then on premise (2). In order to argue against Bogardus's criticism of safety, I aim to show why (2), as it stands, is false. First, let us state the reason we find it true at first glance. To that aim, it is useful to recall one of the definitions of epistemic risk given in chapter 4:

- Unrestricted Belief-focused Risk I: *S*'s belief that *p* is risky$_1$ if and only if *p* is false in at least half the close possible worlds in which *S* believes that *p*.

According to Unrestricted Belief-focused Risk I, an agent who uses a belief-forming method that is as reliable as tossing a coin is at risk of forming false beliefs: the proportion of guesses would be no better than chance. However, in other cases the risk of believing a false proposition is much higher. These are the kind of cases I am interested in. For what I ultimately want to show is that safety is consistent with cases of *high* epistemic risk. We can define (one sense of) high epistemic risk in the following way:

- High Epistemic Risk: *S*'s belief that *p* is at risk of being false if and only if *p* is false in *most* close possible worlds in which *S* believes that *p*.

If we reconsider Atomic Clock, we notice that Smith's belief that it is 8:22 am is epistemically risky in this sense because her belief would be false in very many close possible worlds. In those worlds, the alien isotope decays disrupting the sensor and freezing the clock on the reading "8:22" before Smith looks at it at a time subsequent to 8:22 am. I will argue that, although Smith would easily have formed the false belief that it is 8:22 am, i.e., although her actual known belief is epistemically risky in the sense specified above, her belief is not unsafe (in the sense specified by Safe Belief). In particular, the reason Smith's actual belief counts as safe is that:

- ~(2) in close possible worlds where Smith forms the false belief that it is 8:22 am, she does *not* form the belief via *the same type* of belief-forming method that she uses in @.

~(2) needs justification (one of the goals of this chapter is to provide such a justification). For now, note that ~(2) is *compatible* with the fact that Smith would believe falsely that it is 8:22 am in many close possible worlds, i.e., with the fact that Smith's actual known belief is epistemically risky in the sense specified by High Epistemic Risk. Precisely, –here comes the explanation sought– premise (2) seems true at first glance because Smith's belief is epistemically risky in this sense.

Contra this intuition, I will argue that not every case of epistemic risk is a case of unsafe belief. In other words, I will argue that there are epistemically risky safe beliefs. More specifically, I will defend this claim by arguing that safety is a stringent condition that does not require match between belief and fact across close possible worlds *simpliciter*, but match between belief and fact across close possible worlds in which the agent acquires the relevant belief via the same type of belief-forming method that she uses in @.

First of all, we need to show why modal conditions need to be relativized to methods of belief formation. Recall GRANDMOTHER:[4]

> GRANDMOTHER
>
> A grandmother goes to the hospital to visit her grandson. The grandson's heart is in such a very bad condition that he could easily suffer a deathly cardiac arrest at any moment. If the grandson died, others would tell the grandmother he was alive to spare her upset. The grandmother enters the hospital room and sees that the grandson is alive.[5]

The case illustrates why safety needs to be relativized to belief-forming methods. Intuitively, the grandmother knows that her grandson is alive when she looks at him. Nevertheless, the grandson could easily have died and the grandmother could easily have formed the false belief that he was alive (because others would have told her so). That is, the grandmother's actual belief is epistemically risky in the sense defined by HIGH EPISTEMIC RISK.

However, from the fact that her belief is epistemically risky in the specified sense, it does not follow that it is unsafe. According to SAFE BELIEF, the target belief counts as safe because close possible worlds in which the old woman believes that her grandson is alive via *the same type* of method that she uses in @ (a visual method) are worlds in which the grandson has not yet suffered a cardiac arrest and, consequently, lives. In other words, close possible worlds in which the grandmother believes falsely that her grandson is alive (most close possible worlds) do not make the grandmother's actual belief unsafe because those are worlds in which she forms the target belief by testimony, i.e., by a method of *a different type* to her actual method of belief formation.

It is *prima facie* plausible to maintain that GRANDMOTHER and ATOMIC CLOCK are not structurally equivalent. Contra this intuition, I will argue that they have the same general structure by showing that the relevant beliefs of both agents, although epistemically risky, are not unsafe. More specifically, my reasoning will be as follows: 1) the agents come to know the target propositions via belief-forming methods of certain type in @. 2) In most close possible worlds, those propositions are false, i.e., the target actual beliefs are epistemically risky (as de-

4 See section 4.1.1.

5 I have modified Nozick's original case (Nozick 1981: 179) in such a way that the target proposition could easily be false, as in ATOMIC CLOCK.

fined by High Epistemic Risk).[6] 3) However, this does not mean that they are unsafe, because close possible worlds in which the believed propositions are false are worlds in which the corresponding beliefs are formed via methods of a different type to the methods that the agents use in @.

In general, my strategy to diminish the force of Bogardus's counterexample is roughly that: to show that Atomic Clock is structurally equivalent to several cases of epistemic risk in which the target beliefs are known *and* safe. For that, we need a principled way of individuating methods of belief formation that allows us to give a principled explanation of why a method used by an agent in @ is of a different type to the method that she uses in close possible worlds in which she believes the same proposition falsely. With such a principle, we could give a uniform explanation of all cases structurally equivalent to Atomic Clock and, thus, we would be able to justify ~(2) in a non-question-begging way. It is important to emphasize that the treatment of Atomic Clock must be the same as of the rest of the cases (i.e., it must be principled); otherwise, our reply to Bogardus's counterexample would be *ad hoc*. Desirably, the principle of individuation should also offer the correct diagnosis of paradigmatic cases of unsafe belief. As we will see, an externalist principle of individuation will satisfy all these desiderata.

9.1.2 *Relativization to Methods of Belief Formation*

Mark Alfano (2009) discusses several ways of relativizing Nozick's modal conditions (sensitivity and receptivity) to methods of belief-formation. Here, I will discuss several of the principles that Alfano discusses so as to judge whether they could be used to relativize safety to methods of belief formation. Let m_1 and m_2 be two belief-forming method tokens and M a belief-forming method type. Consider the following principle:

- (R1) m_1 and m_2 are of the same type M if and only if m_1 and m_2 are both cases of vision or olfaction or audition or taction or gustation or testimony or deduction or induction or memory.[7]

(R1) offers the correct diagnosis of Grandmother. In @, the old woman uses a belief-forming method based on vision (she looks at her grandson). In close possible worlds in which her grandson dies (most of them), she uses a belief-forming method based on testimony (others tell her that her grandson is alive). It thus follows from (R1)

6 In Atomic Clock , the cause of epistemic risk is the isotope, which could easily stop the clock. In Grandmother, the source of epistemic risk is the bad condition of the grandson's heart, which could easily collapse.

7 Adapted from Alfano (2009: 279). As Alfano explains, (R1) may need a few more disjuncts, but the idea should be intuitively clear. See Goldman (2009: 80-82) for relevant discussion on this way of individuating belief-forming methods.

that the grandmother's actual belief-forming method is of a different type to the one she uses in most close possible worlds. Since, according to SAFE BELIEF, relevant close possible worlds are only those in which the agent forms the target belief via the same type of method that she uses in @, most close possible worlds are irrelevant to assess whether the grandmother's actual belief is safe. It is safe indeed, as in nearly all (if not all) close possible worlds where her grandson lives (not many of them) she *sees* that he is alive.

Note that (R1) offers an incorrect diagnosis (at least for our purposes) of ATOMIC CLOCK. In particular, it fails to explain why Smith's belief-forming method in @ is of a different type to the method that she uses in most close possible worlds. On the one hand, Smith looks at the working clock in @. Thus, the actual method token is a case of vision. On the other hand, in close possible worlds in which the isotope decays and stops the clock (most of them), Smith looks at the stopped clock and forms the false belief that it is 8:22 am. Therefore, the relevant method tokens in those possible worlds are cases of vision too. It thus follows from (R1) that all methods (in @ and in close possible worlds) are tokens of the same method type. Then, it follows from SAFE BELIEF that Smith's belief is unsafe, as in nearly all (if not all) close possible worlds she believes the same proposition by the same type of method and the proposition is false. Given the safety condition, her actual belief is not knowledge, which contradicts premise (1).

This result is certainly bad for our purposes of defending safety from Bogardus's attack. However, we cannot conclude (on pain of begging the question) that (R1) is unsatisfactory just because it fails to offer the diagnosis of the case that is convenient for our purposes. That is to say, we need to justify the tenability of principles of method individuation independently of Bogardus's case. To that aim, I will put aside, for the moment, ATOMIC CLOCK and Bogardus's contention that unsafe knowledge is possible and I will try to answer to the following question: if one were in the position of the safety theorist, which is the best principle of method individuation that one could adopt, i.e., which principle would allow to account for more cases? To answer this question, we need to take into account a variety of cases and see whether the different principles of method individuation allow the safety theorist to explain them in keeping with our intuitions about whether they are cases of knowledge or not.

To begin with, as Alfano explains, (R1) fails to provide (as far as the purposes of the safety theorist are concerned) the correct diagnosis of the following case:

> REDWOOD
>
> Suppose Scotty sees and correctly identifies a redwood as a tree. Suppose further that there are plants nearby that Scotty would mistake for bonsai trees if he were to see

> them and that Scotty would have seen them if he did not see the redwood. Nevertheless, it seems, Scotty knows that he sees a tree. (Alfano 2009: 276)[8]

If we relativized the safety condition to methods of belief formation individuated according to (R1), we would have to claim that Scotty's belief that the object in front of him is a tree (when he looks at the huge redwood) is unsafe. See why. Presumably, in many close possible worlds Scotty looks at the nearby plants and comes to believe that they are trees (although they are not). Looking at nearby plants is a case of vision. So is looking at the redwood. Thus, (R1) considers all the relevant method tokens of the same type (cases of vision). Then, given SAFE BELIEF, Scotty's belief that there is a tree (when he looks at the redwood) is unsafe and hence, given the safety condition, not knowledge.

We can thus conclude that (R1) is not a good option for the safety theorist. Let us consider another proposal that can be traced back to Goldman (1976: 779- 780) and to Nozick (1981: 184):

- (R2) m_1 and m_2 are of the same type M if and only if m_1 is experientially the same "from the inside" as m_2.

(R2) offers the correct diagnosis of GRANDMOTHER. In particular, it explains why the grandmother's belief-forming method in @ is of a different type to the one she uses in most close possible worlds. In @, the old woman uses a visual method. In close possible worlds in which her grandson dies (most of them), she is told that her grandson is alive. Testimony and vision are not experientially the same from the inside. Therefore, (R2) considers the grandmother's belief-forming method in @ of a different type to the method that she uses in close possible worlds where she forms a false belief. Then, given SAFE BELIEF, her actual known belief is safe.

In addition, (R2) also offers the correct diagnosis of REDWOOD. In particular, it explains why Scotty's actual belief-forming method is of a different type to the methods that he uses in most close possible

8 The case was originally introduced by Alvin Goldman (1976: 779). Epistemologists have paid more attention to DACHSHUND, which is structurally equivalent to REDWOOD (see section 8.1.1). In DACHSHUND, Oscar comes to know that the object in front of him is a dog when he looks at a dachshund despite the fact that he could easily have believed the same proposition when looking at a wolf (hence falsely). Goldman intends to show that the Oscar's inability to discriminate wolfs from dachshunds does not prevent him from successfully identifying dachshunds as dogs: "Oscar's true belief fails to be knowledge if there is an alternative situation in which a non-dog produces the same belief by means of the same, or a very similar, appearance. But the wolf situation is not such an alternative: although it would produce in him the same belief, it would not be by means of the same (or a similar) appearance" (Goldman 1976: 779). By contrast, a situation with dachshund replicas would be such an alternative: a non-dog would produce the same belief by means of the same, or very similar, appearance.

worlds. In @, Scotty has a redwood-like experience. In close possible worlds where he looks at the nearby plants (presumably, most of them), he has bonsai-like experiences. Consequently, it follows from (R2) that his actual method is of a different type to the methods that he uses in close possible worlds where he looks at the nearby plants. In this way, given Safe Belief, his actual known belief counts as safe, because close possible worlds where he uses an experientially equivalent method to form the belief that the object in front of him is a tree are worlds in which the object in front of him is a tree, namely a redwood.

Note that (R2) does not offer the correct diagnosis (at least for our purposes) of Atomic Clock. In particular, it does not explain why Smith's belief-forming method in @, where she looks at the working clock, is of a different type to the methods that she uses in close possible worlds in which she looks at the stopped clock (most close possible worlds). To see this, just note that in @ and in all close possible worlds Smith has equivalent clock-like experiences.

Be that as it may, the safety theorist should reject (R2) for the following reasons. Consider the following case:

> Matrix
>
> Robots dominate the Earth. They hunt humans to use them as sources of bioelectrical and thermal energy. For that task, they use a special pistol that shoots a microchip to the human's head in such a way that it automatically connects her to a simulated reality called the Matrix. Humans do not even notice the change. Unbeknownst to Neo, he has just entered an area full of robots. Robots start shooting microchips to his head but for very improbable reasons they miss. In the meantime, Neo has picked up a stone and has formed the belief that he is holding a stone in his hand. If he had been hit, he would have dropped the stone and he would still have believed (in the Matrix) that he is holding a stone in his hand.

(R2) offers an incorrect diagnosis of Matrix. Intuitively, Neo knows that he is holding a stone in his hand, even if in most close possible worlds he drops the stone and still believes the same proposition (Neo's actual belief is epistemically risky in the sense specified by High Epistemic Risk, but that does not prevent it from being knowledge). Of course, we do not want Safe Belief to imply that Neo's actual belief is unsafe. And yet, this is the outcome if methods of belief formation are individuated as per (R2). For in @ and in all close possible worlds Neo has the experience that he is holding a stone and (R2) consequently implies that all belief-forming methods are of the same type.

Therefore, the safety theorist has good reason to reject (R2). Let us consider now an externalist proposal:

- (R3) m_1 and m_2 are of the same type M if and only if the relevant external factors involved in the target belief's etiology when formed via m_1 are the same as when formed via m_2.[9]

Let me explain how to understand (R3). I follow John Greco (2005: 266) in thinking that a belief's etiology "concerns such things as the history of the belief and the reasons why it is held". Intuitively, relevant external factors involved in a belief's etiology may be perceived objects such as chairs, tables or hands. To get a more accurate idea of how (R3) classifies types of belief-forming methods, the best thing we can do is to consider some examples.[10]

According to (R3), 1-3 are (all other things being equal) pairs of different types of belief-forming methods: 1) believing that there is a dachshund by seeing a dachshund vs. believing the same proposition by seeing a wolf; 2) believing that one is drinking pure, unadulterated water by drinking pure, unadulterated water from a glass vs. believing that one is drinking pure, unadulterated water by drinking water from a glass that has been doctored with undetectable toxins by conniving agents; 3) believing that one was shown n number of flashes after drinking regular orange juice vs. believing that one was shown n number of flashes after drinking a glass of orange juice with a tasteless mind-altering drug. These examples should suffice to get a general impression of how (R3) works: in each instance of belief formation there is some external factor in the history of the belief that differs from the corresponding instance.

Let us analyze how (R3) deals with the cases scrutinized so far. In REDWOOD, MATRIX and GRANDMOTHER the external factors involved in the etiologies of the actual beliefs of Scotty, Neo and the grandmother are different from the external factors involved in the etiologies of the beliefs that they would form in most close possible worlds. Compare the relevant pairs of belief-forming method tokens: 1) believing that there is a tree by seeing a huge redwood (@) vs. believing that there is a tree by seeing a shrub (most close possible worlds); 2) believing that one is holding a stone by seeing a stone in one's hand (@) vs. believing that one is holding a stone by being connected to a simulated reality; 3) believing that one's grandson is alive by seeing him alive (@) vs. believing it through the testimony of someone who has seen him dead (most close possible worlds). (R3), therefore,

9 Goldman (2009: 81) discusses a principle of individuation like (R3): "[A] possible construal of 'bases' would include specific external objects involved in the method of belief acquisition. The basis of belief in the dachshund case might be *seeing the dachshund*, and the basis of belief in the wolf case might be *seeing the wolf*" (instead of talking about methods of belief formation, Goldman talks, more generally, about bases of belief). In addition, Timothy Williamson (2009: 307) seems to assume a principle of individuation like (R3) but, at some point, he also seems to hold something along the lines of my proposal, (R4), see e.g., Williamson (2009: 325 fn.13).

10 The following examples are discussed by Williamson (2009: 307) and are listed by Dani Rabinowitz (2011), a list that I copy (almost literally) in the next paragraph.

rules that the methods of belief formation used by Scotty, Neo and the grandmother in most close possible worlds are different from the methods that they use in @. Their actual true beliefs can in this way count as safe.

Note, in addition, that (R3) does not seem to offer the correct diagnosis (at least for our purposes) of Atomic Clock. In @, Smith looks at a clock whose mechanism is in motion and forms the true belief that it is 8:22 am. In close possible worlds where Smith forms the false belief that it is 8:22 am (most of them), she looks at the same clock, which is stopped, however. Since the most relevant factor of the etiology of the relevant beliefs in all worlds is the clock and its reading "8:22", it does not seem that (R3) should consider Smith's actual method of a different type to the methods that she uses in most close possible worlds.[11]

At any rate, the safety theorist has other reasons to reject (R3). In particular, it fails to offer the correct diagnosis of Fake Barns. Recall the case (a short version):

> Fake Barns
>
> Henry forms the true belief that the object in front of him is a barn. Although the object is a genuine barn, Henry does not know it because the environment is populated with indistinguishable barn replicas that would easily have led him to form false beliefs in the same proposition.

In @, Henry looks at a genuine barn and forms the belief that the object in front of him is a barn. In close possible worlds where he believes the same proposition falsely (most of them), he looks at barn replicas. Since the etiologies of the relevant beliefs in most close possible worlds involve objects (barn facsimiles) that are different to the most relevant object of the etiology of the target belief in @ (a genuine barn), Henry's actual belief-forming method is, according to (R3), of a different type to the methods that he uses in most close possible worlds. However, this result is disastrous for the safety-theorist: if she cannot fix that in close possible worlds Henry acquires his beliefs via the same type of belief-forming method, she is compelled to regard Henry's actual belief as safe, which goes against the standard diagnosis of the case. Therefore, (R3) is useless for the safety theorist.

Let us take stock. The cases discussed thus far show that the safety theorist is in need of a principle of individuation that can guarantee, first, that in Fake Barns (a case where the agent lacks knowledge) the actual belief-forming method is of *the same type* as the methods that the agent uses in close possible worlds where the relevant beliefs

11 Nevertheless, if the decay of the isotope is part of the relevant etiologies, then (R3) does consider Smith's actual method different to the methods used in close possible worlds. It all depends on how fine-grained are the criteria for individuating etiologies.

are false (most close possible worlds). In that way, the agent's actual belief can count as *unsafe*. Second, the principle of individuation must guarantee that in Grandmother, Redwood and Matrix (all of them cases of knowledge) the target belief-forming methods in @ are of *a different type* to the methods that the agents use in close possible worlds where their beliefs are false (most of them). In this way, their actual beliefs can count as *safe* and thus qualify for knowledge.

9.1.3 *An Externalist Principle of Individuation*

In what follows, I will develop an externalist principle of method individuation that will offer a solid and principled justification for ~(2). To give shape to the proposal, we need to introduce a common distinction in epistemology (due to McGinn 1984) between two notions of reliability: *local* and *global* reliability. According to McGinn, a locally reliable belief-forming method is a method that is reliable with respect only to the proposition believed, while a globally reliable belief-forming method is a method that is reliable with respect to a range or a field of propositions.

In my view, the distinction between local and global reliability should be drawn not only in terms of the range of propositions about which a belief-forming method is reliable, but also in terms of the range of circumstances about which it is regarded as reliable. After all, when we estimate the degree of reliability of a belief-forming method we not only locate the propositions but also the circumstances relative to which the ratio of true beliefs is highest.

Accordingly, the set of circumstances with respect to which a belief-forming method is *locally* reliable includes only those in which the target proposition is believed. By contrast, a *globally* reliable belief-forming method is a method that is reliable with respect to a field of propositions and a range of circumstances. Circumstances here are understood as sets of external conditions of multiple environments (e.g., good light conditions) plus the internal conditions of the agent in those environments (e.g., being in good shape, sober, and so on).[12] Fields of propositions may be, for example, propositions about certain color or propositions of the type '*n* is an even number', where '*n*' is replaced by any number.

The fact that two belief-forming method tokens are *globally* reliable to the same degree implies that both of them tend to produce similar cognitive performances concerning the same kind of propositions and circumstances. This is a good indication, but not a sufficient reason (as we will see later), to think that they are the same type of method. For example, when two visual-based methods of belief formation are of the same type their degree of global reliability will be highest and,

12 See Sosa (2010: 465-7) for relevant discussion on this particular way of characterizing circumstances.

plausibly, the same with respect to propositions about, say, the color red (rather than propositions about tones of voice) and with respect to circumstances in which the light conditions are good (rather than, say, circumstances without ambient noise). We can accordingly formulate the following necessary condition, which will be part of our principle of individuation:

- (i) m_1 and m_2 are of the same type M, only if they are globally reliable to the same degree with respect to the same field of propositions and the same range of circumstances.

The notion of knowing via a method was introduced by Nozick (1981) to protect his analysis of knowledge from counterexamples similar to GRANDMOTHER. One might think that there is such a thing as a commonsense notion of method of belief formation. This notion is plausibly captured by a principle like (R1), which individuates methods in terms of the modality employed in forming the belief. If we do not include in our principle of individuation some condition along these lines, we may have to deal with the objection that individuating methods *just* in terms of their global reliability does not take into consideration our ordinary intuitions about belief-forming methods. It would be desirable, therefore, to include such a condition in our principle of individuation.

However, we have seen that (R1) individuates methods in an incorrect way for the safety theorist, as REDWOOD shows. But this does not mean that there is nothing we can preserve from (R1). Note that (R1) says two things: (1) that if m_1 and $m_2$are both cases of vision or olfaction or audition, and so on, they are the same type of belief-forming method and (2) that if m_1 and m_2 are the same type of method, then they are both cases of vision or olfaction or audition, and so on. Cases like REDWOOD prove (1) wrong for the purpose of formulating a successful safety condition, but no case so far considered shows that (2), the necessity claim, is not suitable for that aim. Accordingly, we can include a requirement similar to (2) in our principle of individuation so as to take into account the modality employed in the formation of the belief. More specifically:

- (ii) m_1 and m_2 are of the same type M, only if they are both based on vision or olfaction or audition or taction or gustation or testimony or deduction or induction or memory.

However, the conjunction of conditions (i) and (ii) may not be sufficient to individuate methods of belief formation in a correct manner. Consider REDWOOD. In @, Scotty uses a method token (m_1) that allows him to identify the redwood as a tree. It is a stipulation of the case that m_1 is globally reliable, and also that Scotty's actual circumstances (where there is a redwood within his field of vision) belong to the set of circumstances with respect to which m_1 is globally reliable. In the counterfactual scenario, Scotty looks at the nearby plants (e.g., shrubs) and thinks that they are bonsai trees using method token m_2. As the case is described, nothing prevents us from assuming that m_2 is globally reliable in the same way as m_1, i.e., to the same degree and with respect to the same propositions and circumstances. Accordingly, if the conjunction of (i) and (ii) is all it takes to individuate methods, why should we think that m_1 and m_2 are not the same type of method? After all, they are both based on vision and globally reliable in the same way.

At this point, it is clear that we need to supplement conditions (i) and (ii) with a further requirement. Let me motivate a condition that will turn our principle of individuation into a distinctly externalist proposal. Quite often, the expressions 'belief-forming method' and 'belief-forming process' are interchangeably used in the epistemological literature. Nevertheless, there are subtle differences between the notions to which these expressions refer. By 'belief-forming process' is usually meant, as Goldman (1986: 297) puts it, "a *basic* psychological process, something inherent in the fundamental architecture of the human cognitive system". A basic psychological process may certainly be part of a belief-forming method, but the notion of method is broader, as I will argue next. Crucially, methods of belief formation are, as Nozick (1981) originally introduced them, *ways of believing*.

In general, when we individuate ways of φ-ing, we not only take into account the abilities needed to φ, but also conditions of the circumstances that are relevant to the deployment of such abilities. For example, a way of climbing the Mount Everest is to make such-and-such moves without supplemental oxygen via the southeast ridge under such-and-such climate conditions. Ways of believing are no exception. In the same way as expert mountaineers classify ways of climbing in terms of the use (or not) of supplemental oxygen, the route followed or the season of the year, we are allowed to appeal to conditions of the circumstances that are relevant to belief formation in order to individuate ways of believing. In particular, relevant conditions are those with respect to which a way or method (whether based on one or another modality) is (to a certain degree) globally reliable.

Accordingly, the third condition of our principle of individuation can be stated as follows (let m_1 be the target belief-forming method):

- (iii) m_1 and m_2 are of the same type M, only if the circumstances in which the target belief is formed via m_2 are in the set of circumstances with respect to which m_1 is globally reliable.

Conditions (i), (ii) and (iii) make up a plausible externalist principle of individuation:

- (R4) m_1 and m_2 are of the same type M if and only if (i) m_1 and m_2 are globally reliable to the same degree with respect

> to the same field of propositions and the same range of circumstances; (ii) they are both based on vision or olfaction or audition or taction or gustation or testimony or deduction or induction or memory; (iii) the circumstances in which the target belief is formed via m_2 are in the set of circumstances with respect to which m_1 is globally reliable.

Let us analyze Redwood again. The method that Scotty uses in @, m_1, is some vision-based globally reliable method that enables him to identify redwoods as trees. Scotty's actual circumstances belong to the set of circumstances with respect to which m_1 is globally reliable. In this big set of circumstances we can find circumstances where the light conditions are good, where there are (within Scotty's field of vision) objects with trunks, branches and leaves with certain characteristics such as having such-and-such shape or being at a considerable distance from the ground, as well as many other relevant features. In the counterfactual scenario where Scotty uses method token m_2, there are (within his field of vision) objects with trunks, branches and leaves at a little distance from the ground, and so on.

The distance of the leaves from the ground, the width of the trunk as well as many other relevant factors of the circumstances of m_2 differ from the factors that shape the kind of circumstances with respect to which m_1 is globally reliable. Thus, condition (iii) is not met, which means that m_1 and m_2 are not the same type of method. Consequently, close possible worlds in which Scotty believes that there is a tree when he looks at, say, a shrub are not relevant to evaluate whether his actual true belief is safe. Since in close possible worlds where he looks at the redwood (presumably, few close possible worlds), he correctly identifies it as a tree, his actual true belief is safe, according to Safe Belief.

Let us consider now Grandmother. In this case, condition (ii) is not met. In @, the old woman uses a vision-based belief-forming method. By contrast, in close possible worlds in which her grandson dies (most of them), she uses a testimony-based method. Thus, her actual method is of a different type to the method that she uses in most close possible worlds, which means that close possible worlds in which she believes that his grandson is alive falsely are not relevant to evaluate whether her actual true belief is safe. Since in close possible worlds where her grandson is still alive (few close possible worlds), she comes to believe that he is alive by looking at him, her actual true belief is safe, according to Safe Belief.

In Matrix, condition (ii) is not met either. In @, Neo forms the true belief that he is holding a stone in his hand by looking at his hands. Thus, he uses a vision-based belief-forming method. In close possible worlds where he forms the belief that he is holding a stone in his hand and the belief is false (most of them), he is connected to the Matrix and a supercomputer feeds him false information. Whatever this method is, it is not based on vision, and, therefore, it is of a different type to Neo's actual method.

One could modify the case in such a way that the supercomputer fed Neo false information via some vision-based method, so that condition (ii) would be met. In that scenario, however, it is dubious that condition (iii) would hold. For circumstances in which Neo comes to believe that he is holding a stone being plugged in to the Matrix are not the type of circumstances with respect to which his actual vision-based method is globally reliable. After all, when we estimate the *global* reliability of an ordinary visual method (like Neo's actual method), we do not take into consideration circumstances in which one is plugged in to a supercomputer. Therefore, close possible worlds in which Neo is connected to the Matrix (most of them) are not relevant to assess whether his actual true belief is safe. Since in close possible worlds where he is not in the Matrix (few of them), he looks at his hand and forms the true belief that he is holding a stone, his actual belief is safe, according to SAFE BELIEF.[13]

Consider FAKE BARNS. The standard diagnosis of the case says that Henry does *not* know that the object in front of him is a barn, although he is looking at a genuine one. (R4) explains why Henry's actual method is of the *same* type as the method of belief formation that he uses in close possible worlds where he looks at barn replicas. To see this, I will consider each condition separately. Let m_1 be Henry's actual method and m_2 the method that he uses in close possible worlds in which he believes that there is a barn in front of him when looking at a barn replica.

Condition (ii) obviously holds, since both m_1 and m_2 are vision-based methods. Condition (i) is also met. The fact that neither m_1 nor m_2 are *locally* reliable (reliable relative to the specific circumstances where they are used) does not prevent them from being globally reliable to the same degree and with respect to the same propositions and circumstances. In particular, m_1 and m_2 are *globally* reliable methods for discriminating barns from other buildings such as houses, garages, airport hangars or skyscrapers (i.e., with respect to propositions and circumstances associated to these items). As the case is described, there is no reason to think that Henry does not have that ability and that he does not retain it when he fails to know in the Fake Barn Country that the object in front of him is a barn, viz., when his ability is not *locally* reliable.

On the other hand, circumstances where there is an object with the characteristics of a barn within his field of vision, where the light conditions are such-and-such, where Henry is at distance n, and so

13 As a general rule, if (ii) is not met, then (i) and (iii) are not met either, i.e., if m_1 and m_2 are not cases of the same modality, then neither m_1 and m_2 are globally reliable to the same degree with respect to the same field of propositions and the same range of circumstances, nor the circumstances in which the target belief is formed via m_2 are in the set of circumstances with respect to which m_1 is globally reliable.

on, belong to the set of circumstances with respect to which the ability is globally reliable. His actual circumstances, where he uses m_1, are certainly in that set (he is in front of a prototypical barn, the light and distance conditions are good), and so is the counterfactual scenario where he uses m_2: the light and distance conditions are equally good and he is in front of a faithful barn replica, i.e., in front of an object with all the features that typically allow him to differentiate barns from other buildings reliably. Consequently, condition (iii) holds.

Since all conditions of (R4) are met, it follows that m_1 and m_2 are the same type of belief-forming method. In addition, since in most close possible worlds where Henry believes that the object in front of him is a barn via the same type of method his belief is false, his actual belief is unsafe, according to SAFE BELIEF, and not knowledge, given the safety condition, which agrees with the standard diagnosis of the case.

Let us consider now the alleged counterexample to safety, ATOMIC CLOCK. (R4) explains why Smith's actual belief-forming method is of a different type to the method that she uses in most close possible worlds. To see this, I will analyze each condition separately. Let us start with condition (ii). In @, Smith looks at the working clock and forms the true belief that it is 8:22 am. In most close possible worlds, she looks at the stopped clock and believes the same proposition falsely. Both belief-forming method tokens are vision-based, therefore condition (ii) holds. Does condition (i) hold? That is, are method token m_1 (the method of forming beliefs by looking at the working clock) and method token m_2 (the method of forming beliefs by looking at the stopped clock) globally reliable to the same degree with respect to the same field of propositions and the same ranges of circumstances?

Intuitively, the answer is negative. Let us see this in more detail. Plausibly, m_1 and m_2 are relativized to the same field of propositions, namely propositions about the time (rather than, say, propositions about colors). Are m_1 and m_2 globally reliable to the same degree with respect to the proposition that it is 8:22 am? If Smith used m_1 in a range of different situations to form the belief that it is 8:22 am, her belief would be true most of times. Obviously, the same cannot be said about m_2. Therefore, condition (i) does not hold.

Note, in addition, that m_1 has such a high degree of reliability only relative to certain set of circumstances. The kind of circumstances contained in this set are, for example, circumstances in which the light conditions are good, in which the screen of the clock is not covered by an opaque layer of dust, in which the clock is not under strong magnetic influence or under water, in which the voltage level of the battery is not low (with low batteries clocks slow down) and, more importantly, circumstances in which a radioactive isotope has not decayed in the vicinity causing the clock to stop. In the circumstances in which Smith uses m_2, a radioactive isotope has decayed disrupting the clock's sensor and stopping the clock. Therefore, those circumstances are not in the set of circumstances with respect to which m_1, Smith's actual method, is globally reliable. Therefore, condition (iii) does not hold.

Note that it is plausible to maintain that m_1 and m_2 are not the same type of methods because we have conceived methods of belief formation as *ways of believing*. Thus, in the same manner as climbing the Mount Everest by using a functioning oxygen bottle is a different *way of climbing* to climbing the mountain with a broken oxygen bottle, coming to believe that it is 8:22 am by looking at a perfectly working clock is a different *way of believing* to coming to believe that it is 8:22 am by looking at a stopped clock (no matter that the clock or the oxygen bottle are the same objects when they work and when they do not).[14]

In sum, m_1 and m_2 are different types of methods of belief formation, according to (R4). In this way, close possible worlds in which Smith forms the belief that it is 8:22 am by looking at the stopped clock are not relevant to assess whether her actual true belief is safe. Since in close possible worlds in which she forms the same belief via the same type of method (few close possible worlds), the believed proposition is true, Smith's actual true belief is safe, according to Safe Belief. Therefore, all Atomic Clock shows is that safety needs to be relativized to methods of belief formation. We can thus conclude that Bogardus has not made a case for the claim that unsafe knowledge is possible.

9.1.4 *Unsafe, Safe and Super-Safe Belief*

I will conclude with some general remarks on the notions of safety and epistemic risk with the hope that they help to dissipate a common misunderstanding about the kind of cases that are considered

14 One might be concerned about the apparent structural similarity between Atomic Clock and Fake Barns. In Fake Barns, the prototypical features that allow Henry to identify an object as a barn are shared both by genuine and fake barns. This is partly the reason why circumstances with barn replicas belong to the set of circumstances with respect to which Henry's actual visual method is globally reliable (to belong to that set, the light conditions of the circumstances must be good as well, the distance must be appropriate, and so on). In Atomic Clock, one might argue, the prototypical features that allow Smith to read clocks are shared by the working and by the stopped clock (both read 8:22 am). Should then circumstances in which the clock is stopped be part of the set of circumstances with respect to which Smith's actual method is globally reliable? The answer is negative. Smith's actual method is truth-conducive because the clock is a *reliable indicator* of the time (according to Bogardus, it is the world's most accurate clock). Consequently, circumstances in which the clock is stopped are not the kind of circumstances with respect to which Smith's actual method is globally truth-conducive (or reliable).

potential counterexamples to the necessity of the epistemic condition. Consider the following case of knowledge:

> HANDS
>
> George is in his room. There are good light conditions. Nothing is epistemically amiss. He looks at his hands and forms the belief that he has hands.

Intuitively, George knows that he has hands. On the one hand, his belief is safe (as defined by SAFE BELIEF), because in nearly all (if not all) close possible worlds in which he forms the same belief via the same type of method that he uses in @, the belief is true. On the other hand, George's belief is *not* epistemically risky (as defined by UNRESTRICTED BELIEF-FOCUSED RISK I), because in most close possible worlds in which George believes that he has hands, he has hands. Call a belief *super-safe* just in case it is (1) safe and (2) not epistemically risky (in the specified senses of safety and epistemic risk).[15]

The target beliefs in GRANDMOTHER, REDWOOD, MATRIX and ATOMIC CLOCK all amount to knowledge. As we have seen, they are safe because in nearly all (if not all) close possible worlds in which the agents come to believe the same propositions via the same type of methods that they use in @, the propositions are true. However, although safe, the beliefs in question are not super-safe, because they are epistemically risky: in most close possible worlds in which the agents come to believe the same propositions (no matter by which *type* of belief-forming method), the propositions are false.

The following table compares the different modal profiles of unsafe, safe and super-safe beliefs ('@' is the actual world and 'W_n' close possible worlds):

	@	W_1	W_2	W_3	W_4
UNSAFETY	p	~p	~p	~p	~p
	S uses M	**S uses M**	**S uses M**	**S uses M**	**S uses M**
	Bp	Bp	Bp	Bp	Bp
SAFETY	p	~p	~p	p	~p
	S uses M	~(S uses M)	~(S uses M)	**S uses M**	~(S uses M)
	Bp	Bp	Bp	Bp	Bp
SUPER-SAFETY	p	p	p	p	p
	S uses M	**S uses M**	**S uses M**	**S uses M**	**S uses M**
	Bp	Bp	Bp	Bp	Bp

15 This definition of super-safe belief should be extended to the other senses of epistemic risk. For simplicity, I define it in terms of lack of risk of believing a false proposition.

I surmise that the reason some epistemologists have thought that the safety condition is not necessary for knowledge is that they have mistaken super-safety for safety, in such a way that they have proposed cases of knowledge with high levels of epistemic risk involved, thinking them to be obvious counterexamples to safety, when in reality they are only counterexamples to the view that super-safety is necessary for knowledge. However, while super-safety is very desirable from an epistemic point of view (agents with super-safe beliefs are in modally robust epistemic positions), only safety is considered a necessary condition for knowledge.

The tendency to mistake super-safety for safety is due to a misunderstanding about how to individuate methods of belief formation. In particular, it is thought that two belief-forming method tokens m_1 and m_2 are of the same type if they are both cases of vision or audition or deduction, and so on. However, I have shown that this way of individuating methods leads to a wrong diagnosis of certain cases (it is the other direction that is correct: m_1 and m_2 are of the same type only if they instantiate the same modality). The core of the mistake lies in thinking that if m_1 and m_2 are instances of the same modality, they will always be the same type of method regardless of the circumstances in which they are used. However, I have shown that if m_1 is globally reliable to certain degree, m_2 is not the same type of method unless the circumstances in which m_2 is used are contained in the set of circumstances with respect to which m1 is globally reliable.

There is something compelling about this thought. For example, Terry the taxi driver might be a very reliable way for you to go home, unless Terry is completely drunk, in which case Terry's taxi is a terribly unreliable way of getting home. The degrees of reliability here are so different that we judge that these are different ways of getting home, even though one is also tempted to judge that they are instances of the same type of method (Terry's taxi) used in different circumstances (driving sober vs. driving drunk). Although in some conversational contexts it might be acceptable to claim that driving drunk is the same way of driving a car as driving sober, such a coarse-grained individuation of ways of φ-ing is not adequate in general. The fact that you take Terry's taxi (rather than, say, Larry's) is only a necessary condition for individuating your way of getting home.

The same applies to the epistemic case: it is a mistake to think that if m_1 and m_2 include the use of the same epistemic device (e.g., a clock or a thermometer), then they are the same type of belief-forming method regardless of the circumstances in which the device is used. In the same manner as shooting with a bow whose string is tight is not the same way of shooting as shooting with a bow whose string is completely loose, or in the same manner as playing a guitar that is correctly tuned is not the same way of playing as playing with an out of tune guitar, believing that p by means of a properly functioning epistemic device is not the same way of believing that *p* as coming to believe that *p* by means of a broken epistemic device.[16] Accordingly, in the same way as the success of a shot is safe even though the archer could easily have lost her method of shooting (e.g., even though her bow could easily have broken), the success of a belief counts as safe even though the agent could easily have lost her method of belief formation (e.g., even though the epistemic device used could easily have broken).

9.2 SECURITY

In chapter 6, I introduced Sosa's AAA performance-assessment structure according to which performances can be assessed along three dimensions: *accuracy* (accurate performances attain their aims, i.e., they are successful performances), *adroitness* (adroit performances are the product of competence or ability) and *aptness* (apt performances are successful because of competence or ability, i.e., they are accurate because of adroit). In the final section of the dissertation, I propose to extend Sosa's AAA performance-assessment structure with another normative property: *security.*

9.2.1 *Secure and Super-Secure Performances*

The *rough* idea is that *secure performances* are apt performances that could not easily have been inapt. But this is only a rough idea because one thing is what I call the property of being secure and another thing is the property that I call *super-security*. Here are the definitions of secure and super-secure performances:

- Secure Performance*:* *S*'s performance in the actual world @ is *secure* if and only if (i) *S*'s performance is apt in @ and (ii) in nearly all, if not all, close possible worlds in which *S* performs *in the same way* as in @ (i.e., using the same type of method), *S*'s performance is apt.

- Super-Secure Performance*:* *S*'s performance in the actual world @ is *super-secure* if and only if *S*'s performance is apt in @ and in nearly all, if not all, close possible worlds.

As we see, what differentiates security from super-security is the relativization to ways of performing. A secure performance is not an apt performance that would be apt in all close possible worlds *simpliciter*, but an apt performance that would be apt in close possible worlds *in which the agent performs in the same way as in the actual world*. Let me give an example to appreciate the difference better:

16 The more specialized the context is (e.g., archery and musical competitions, philosophical contexts), the more salient and fine-grained these differences are.

MAGNETIC ARROW II

> Today is training day. Artemis, the great archer, feels very sleepy and constantly yawns. Yesterday was a long night. She normally fails when she is sleepy. She gets ready to shoot. During the process, she always eat a candy. She takes one from a bag full of candies. Unbeknownst to Artemis, the candy she has taken is a caffeine pill. She suddenly becomes alert, stops yawning, takes aim, shots and hits the bull's eye with great skill. Unbeknownst to Artemis, the target is metallic and her arrow has a *deactivated* strong electromagnet hidden at the tip. A hidden agent has the power of activating the strong electromagnet, something that she would have done if Artemis showed ostensible signs that her shot was not going to be successful (e.g., if she aimed wrongly or if she showed ostensible signs of weakness or sleepiness). Before shooting, the hidden agent knew that Artemis had ingested a caffeine pill and because she also knew that Artemis is a great archer when alert, she has not activated the electromagnet.

Artemis's shot is apt: it is successful because of competent. It is safe as well: in close possible worlds in which Artemis shoots *in the same way as in the actual world* (not many close possible worlds), her shot hits the target. Artemis's shot is super-safe: in most close possible worlds her shot is successful (even though she normally fails when sleepy, in close possible worlds in which she does not take the caffeine pill (most close possible worlds) the hidden agent activates the electromagnet and the arrow reaches the target). Artemis's shot is also secure: in close possible worlds in which she shoots in the same way as in the actual world (not many), her shot is apt. To be clear: shooting sleepy is not the same way of shooting as shooting alert (internal factor); shooting an arrow to a magnetic target that has an activated electromagnet at the tip is not the same way of shooting as shooting an arrow that has a deactivated electromagnet at the tip (external factor). Finally, Artemis's shot *is not super-secure*: in most close possible worlds her shot is not apt, viz., in most close possible worlds it is successful, but does not manifest archery competence, because in most close possible worlds the hidden agent activates the electromagnet, which makes the arrow hit the target.

9.2.2 *Secure and Super-Secure Beliefs*

Security and super-security, which are properties of performances, can be applied to cognitive performances. In particular, we can distinguish between *secure* and *super-secure belief*:

- SECURE BELIEF: S's belief that p formed in the actual world @ via a belief-forming method of type M is secure if and only if (i) it is *apt* in @ and (ii) in nearly all, if not all, close possible worlds in which S forms the belief that p via a belief-forming method of type M, that belief is *apt*.

- SUPER-SECURE BELIEF: S's belief that p is super-secure if and only if (i) it is *apt* in @ and (ii) in nearly all, if not all, close possible worlds in which S forms the belief that p, that belief is *apt*.

The principle of method individuation is, as in the case of safety, (R4). Since aptness is factive and, consequently, security entails safety, the range of possible worlds that is relevant to evaluate whether a belief is secure is the same range of possible worlds that is relevant to evaluate whether a belief is safe. Analogously, the range of possible worlds that is relevant to evaluate whether a belief is super-secure is the same range of possible worlds that is relevant to evaluate whether a belief is super-safe.

9.3 THE CONTROL THEORY OF KNOWLEDGE RE-REVISITED

If a cognitive success is secure, then: it is (1) successful (accurate) and (2) competent (adroit) and therefore (3) arises out of a cognitive ability that is cognitively integrated; in addition, it is (4) apt and (5) safe. (1)-(5) are all necessary for epistemic control. We can now complete the account of epistemic control in the following simple way:

- EPISTEMIC CONTROL: S controls her cognitive performance if only if her cognitive success is secure.

A complete cognitive achievement is a cognitive success that results from a cognitive performance over which the agent has epistemic control. Knowledge is a complete cognitive achievement. Knowledge is cognitive success that results from an epistemically controlled cognitive performance. Knowledge is secure belief:

- THE CONTROL THEORY OF KNOWLEDGE (CTK): S knows that p if and only if S's belief that p is secure.

CTK delivers a correct diagnosis of all the cases discussed so far. For one thing, it is able to explain all cases that safety and aptness can explain separately or in conjunction: cases of knowledge such as perceptual, testimonial, inductive or inferential knowledge, cases of intervening luck (e.g., Gettier's cases) and of environmental luck (e.g., FAKE BARNS), cases in which the direction of fit of the target belief is the wrong one (e.g., TEMP and TEMPY) and cases of veritic fortune (e.g., ACTOR). For another thing, CTK can explain cases in which beliefs are safe, apt and yet not-known (FAKE BARNS II and III). In FAKE BARNS II, for example, close possible worlds in which Henry looks at a fake barn, i.e., cases in which he forms his belief in the same way as in the actual world according to (R4), are worlds in which his belief that some object in the field is not apt (it is not successful because of visual competence). In FAKE BARNS III, close possible worlds in which Henry looks at a fake barn, forms the false belief that there is a barn before his eyes and infers the true belief that some object in the field is a barn are worlds in which Henry's acquisition of evidence for the premise belief is not apt and hence the inferred true belief does not manifest competence. Finally, CTK is much more specific than Turri's account of knowledge in terms of amplitude: it offers a principle of method individuation that explains how and when safety and security are to be applied. In addition, it is better motivated than Turri's account: it does not explain knowledge as an overwhelming cognitive achievement but, simply, as a controlled one. In sum, CTK seems a better theory of knowledge.

9.4 SUMMARY

In section 9.1, I have defended the safety principle from a recent counterexample to its necessity for knowledge by Tomas Bogardus. I have shown that the case that Bogardus proposes is no counterexample, but just corroborates the well-known requirement that modal principles must be relativized to methods of belief formation. In responding to the counterexample, the analysis of several cases and several principles of method individuation has served to construct an externalist principle of method individuation that not only shows that the target known belief of Bogardus's case is safe, but that it also delivers the correct judgment about all the cases analyzed (unlike the other principles of method individuation considered).

The idea has been to conceive methods of belief formation as ways of believing (as Nozick originally conceived them). In general, I have explained, two ways of φ-ing w_1 and w_2 are of the same kind only if they are reliable to the same degree with respect to the same set of circumstances. In addition, if the circumstances in which one actually φ-es are not in that set (i.e., are not the type of circumstances with respect to which w_1 and w_2 are reliable to the same degree), then one's actual way of φ-ing is not of the same type as w_1 and w_2. In brief, relevant factors of the circumstances individuate the way we do things. I have applied this way of seeing things to the epistemic case. Two belief-forming method tokens (or ways of believing) m_1 and m_2 are of the same type only if they are globally reliable to the same degree with respect to the same field of propositions and the same range of circumstances and only if the circumstances in which the a belief is formed via m_2 are in the set of circumstances with respect to which m_1 is globally reliable (and *vice versa*). They are of the same type if, in addition, they are both instances of the same kind of knowledge source (vision, audition, memory, and so on).

At the end of section 9.1, I have offered a diagnosis of why some epistemologists have thought that safety is not necessary for knowledge. They mistake, I have argued, a condition that I call super-safety for the safety condition. If super-safety does not hold, then one fails epistemically in the unrestricted class of close possible worlds, whereas if safety is not satisfied, one fails epistemically in the class of close possible worlds in which one comes to form beliefs in the same way as in the actual world.

In section 9.2, I have extended Sosa's performance-assessment model with a normative property that I have called security. A secure performance is an apt performance that would not easily be inapt were the agent to perform in the same way as she actually does. I have also made a distinction between security and another normative property that I have called super-security. As in the case of safety and super-safety, if super-security does not hold, then one's φ-ing is inapt in the unrestricted class of possible worlds, whereas if security does not hold, then one's φ-ing is inapt only in the class of close possible worlds in which one φ-es in the same way as in the actual world. Finally, I have defined the notion of epistemic control in terms of secure cognitive performance, and consistently with what I have argued in previous chapters, I have defined the notion of cognitive achievement in terms of epistemic control and knowledge as a cognitive achievement. The conclusion has been that knowledge is secure belief, or so says the control theory of knowledge.

CONCLUSIONS

Good conclusions should put together all the pieces of the puzzle in order to help the reader not to miss the forest for the trees. The standard way of doing this is to (1) summarize the contents of the work, (2) highlight its most original and significant contributions to the issues of the area, (3) state the limitations of the research done and (4) indicate future lines of research. Here I will not summarize the contents of the dissertation (a general summary of the contents can be found in the preface and more exhaustive summaries can be found at the end of each chapter). Rather, I will describe the intellectual process behind the project in a way that will cover points (2), (3) and (4).

When I started designing the structure of this project, I suspected (but did not know yet) what my ultimate goal was: to give an account of the nature of propositional knowledge and, more specifically, to give a definition of the concept of knowledge. Careful examination of some of the writings of Robert Nozick and Ernest Sosa had lead me to the firm conviction that a belief must bear a modal relation with a fact so as to constitute knowledge of that fact. And I had some intuitions about how to implement that idea, but it was not entirely clear to me how to turn those intuitions into a full-fledged theory of knowledge, and worse, I was a bit skeptical that anything like that could ever be accomplished.

To begin with, I was well aware that one of most difficult tasks in philosophy (if not the most difficult one) is to develop your own view; moreover, to develop it in an original enough way so that it can properly count as an alternative view to current views on the same thing. But that thought, which is inherent to the state of being an early PhD student, only encourages one to be more avid in pursuing and developing original ideas. My main concern rather came from the recent history of epistemology: I had the feeling that the history of the analysis of knowledge since Gettier's seminal paper has been a story of systematic failure. The more post-Gettier papers I read on the topic, the more I was convinced that the project was doomed to failure. It seemed like there were a lot of if's saying a lot of things about the nature of knowledge none of them conclusively capturing what knowledge really is. Perhaps the project of analyzing knowledge was a nonstarter, I thought.

However, I simply did not share the intuitions of alternative approaches to knowledge, such as the Oxonian treatment of knowledge as a primitive notion, one that cannot or should not be analyzed, or the subtle shift by epistemological contextualists from the analysis of the concept of knowledge to the analysis of knowledge sentences. All these facts together made me think that, perhaps, it was easier to start investigating what is *not* knowledge rather than what knowledge is. The general guideline, as I put it in an initial thesis proposal, was this: in order to analyze the nature of certain thing it is sometimes simpler to focus on cases in which the thing to be analyzed does not obtain and then try to identify some factor shared by all the target cases.

Duncan Pritchard's seminal monograph on epistemic luck was particularly illuminating at this stage. According to Pritchard, a lot of cases of true belief that is not knowledge can be explained by appealing to the fact that they instantiate a specific and definable phenomenon: veritic epistemic luck. I found Pritchard's explanation elegant and his definition of veritic luck intuitive and easy to apply, and for a while I had the impression of having found what I was looking for: a factor shared by a lot of cases of not-known true belief that could explain, precisely, why those beliefs are not knowledge. I thought then that it was worthy to spend some time delving into the topic of epistemic luck more deeply.

Although Pritchard's book was of tremendous help in this regard, I had the increasing impression that Pritchard was telling only part of the whole story, that he was presenting things in a too easy way. To begin with, his modal definition of veritic luck seemed to fit *too* well with his safety condition for knowledge to the extent that a non-lucky true belief is basically a safe belief according to Pritchard's definitions. This made my suspect that something was being overlooked in the explanation. It seemed that by conceiving veritic luck just in terms of modally fragile cognitive success Pritchard was just paving the way for the introduction of a safety-based epistemology, in such a way that the only theory of knowledge that could ever properly deserve the label of anti-luck epistemology was his own theory (or similar ones). What about simple reliabilism and virtue epistemology?, I thought, do not they deserve to be called anti-luck epistemologies? After all, they aim at solving the Gettier problem and nearly all (if not all) Gettier-style cases instantiate, according to Pritchard, veritic luck. All these unsettling doubts urged me to take a step back.

The issue deserved a fresh start. I carefully examined the general literature on the concept of luck and investigated what had been said in the literature on moral luck and free will. I realized a simple fact: almost all commentators accepted that a necessary condition for luck was that the agent must lack control over the lucky event. Why had Pritchard dropped so quickly a lack of control condition in his explanation of epistemic luck? If epistemic luck is just a type of luck, and a lack of control condition is necessary for the latter, why an epistemic analogue of the lack of control condition cannot be also necessary for the former? The discussion on the concept of knowledge needed to be bracketed and a thorough examination of the concept of luck was in order.

I identified the three types of conditions that are typically considered necessary for luck (chance conditions, significance conditions and lack of control conditions) and I decided to analyze each condition separately. I searched the literature for all the relevant different formulations of each condition and I evaluated their plausibility. As I got more acquainted with the topic, a simple but powerful idea started to bloom in my mind: luck is a form of risk. In particular, I thought, luck involves two types of risk: the risk that the lucky event had of not occurring and the risk at which the lucky agent is with respect to the significant lucky event. Subsequently, I came to think that the most plausible way of defining the event-focused sense of risk was in modal terms and that the agent-focused sense of risk was to be cashed out in terms of lack of control.

Everything seemed to fit nicely. On the one hand, this way of conceptualizing luck was compatible with the way others had thought about the concept of luck. On the other hand, it delivered the correct judgment of paradigmatic cases of luck, such as fair lottery wins. But more importantly, by thinking luck as a form of risk I was in a position to take into account a range of cases in which the occurrence of the relevant events are lucky for the agents in question but in which there was no risk that those events failed to obtain. At that point, I thought, there must be a second sense of the terms the terms 'luck' and 'fortune' that is overlooked by our ordinary use of the terms. This is how I came to distinguish fortune from luck in a technical sense, a distinction that, as we have seen, has been very fruitful when applied to the epistemic case.

The fact that the lack of control condition was necessary for both luck and fortune made me seriously take into consideration the hypothesis that knowledge requires something like epistemic control. But here again I encountered the same problem that I encountered with the concept of luck: what is control, in general? Once again, I searched the philosophical and non-philosophical literature for uses of the term 'control', but I barely found general definitions that could serve my purposes.

Only Daniel Dennett's books on free will and a couple of engineering textbooks proved to be useful. Dennett in particular distinguishes a type of control that I subsequently called effective control, which involves causal influence (or a determination relation) between the controller and the controllee. After having thought for while about the way we apply the term 'control' in ordinary discourse, I realized that there is another sense of the term, which I called tracking control. In addition, the engineering textbooks helped me to learn about feedback, feedforward and open-loop systems.

It was time to apply these findings to the epistemic case. The first step was to conceive epistemic luck as a form of epistemic risk, in keeping with my general discussion on luck. Accordingly, I distinguished two forms of epistemic risk: a belief-focused sense and an agent-focused sense. I defined the former in modal terms and the latter in terms of lack of (a not yet developed notion of) epistemic control. Crucially, I distinguished between veritic luck and veritic fortune, which allowed me to give the long waited explanation of what knowledge is not: knowledge is not false belief and knowledge is not true belief that is veritically lucky or veritically fortunate.

Once the overwhelming problem of luck (and fortune) in epistemology was properly understood, I was in a position to give a positive account of knowledge. My hypothesis was that knowledge requires epistemic control, but I did not know yet how to make sense of the notion. At that point, the general account of control given before became very useful. On the one hand, Robert Nozick's metaphor of knowledge as a a matter of tracking the truth seemed a direct application to the epistemic case of my notion of tracking control. In this way, I argued that modal conditions for knowledge such as safety or sensitivity capture the modal dimension of the notion of epistemic control. On the other hand, the conditions of the definition of effective control were the same as the conditions that virtue epistemologists had already used to assess performances (in general) from a normative point of view and that they had subsequently applied to cognitive performances. In this way, I could make sense of the agential dimension of the notion of epistemic control. Finally, I also resorted to the work of Pamela Hieronymi on attitudes to make more complete the account of the notion of epistemic control in view of the predictable reluctance that some epistemologists would express when seeing the term 'control' before the term 'belief'.

In any case, I had the unsettling feeling that my argument for the view that knowledge requires epistemic control could not just come from the direct application of two the senses of the notion of control to the epistemic case. I needed an argument that could account for the normativity of knowledge. This is were the notion of achievement became relevant. What was initially meant to be a section became a chapter. Achievements were conceived as successes over which agents have control and a special type of achievements were distinguished: complete achievements. Knowledge was conceived then as a complete cognitive achievement, one that is indisputable from any dimension of epistemic normative assessment. The remaining work was just mere refinement of the view I had called the control theory of knowledge: a complete cognitive achievement is a secure belief. I had finally found my own view, one that can properly count as an alternative view to current views on the same thing: knowledge.

I realize now that the completion of this project has required to tackle many issues in a sketchy manner and to put some problems aside. To name a few: lottery cases, the question of voluntariness of belief and inferential knowledge deserve more careful analyses. On the other hand, the real missing problem has been skepticism. Understandable limitations of space and time explain these limitations of the text. In return, more careful analysis of these problems and issues might constitute promising lines of research, in addition to the work of reorganizing the contents of the dissertation in the form of research papers. Outside epistemology, the distinctions between luck and fortune and tracking and effective control might find valuable applications in the debates on moral luck and free will. Finally, I have lately been playing around with the idea that, in the same way as a significance condition is necessary for luck, an epistemic significance condition is also necessary for epistemic luck, and that this condition might help to support or perhaps to reject pragmatic encroachment (the view that knowledge is partially determined by non-epistemic factors such as factors concerning action). For the moment, all this possible research has only resulted in merely possible papers about actual philosophical problems. I hope to bring them to actuality soon.

BIBLIOGRAPHY

Adams, F. & Clarke, M. (2005). Resurrecting the tracking theories. *Australasian Journal of Philosophy*, 83, 207-221.

Adler, J. (2005). Reliabilist justification (or knowledge) as a good truth-ratio. *Pacific Philosophical Quarterly*, 86, 445-458.

Alfano, M. (2009). Sensitivity theory and the individuation of belief-formation methods. *Erkenntnis*, 70: 271-281.

Alston, W. (2005). *Beyond "Justification": dimensions of epistemic evaluation*. Cornell University Press.

Åström, K. J. Murray, R. M. (2008). *Feedback systems: An introduction for scientists and engineers*. Princenton University Press.

Audi, R. (1993). *The structure of justification*. Cambridge University Press.

Baier, A. (1986). Trust and antitrust. *Ethics*, 96, 231-260.

Balaguer, M. (2010). *Free will as an open scientific problem*. MIT Press.

Ballantyne, N. (2012). Luck and interests. *Synthese*, 185, 319-334.

Barker, S. F. (1987). Conditionals and skepticism. In Luper-Foy, S. (Ed.): *The possibility of knowledge: Nozick and his Critics*. Rowman & Littlefield.

Baumann, P. (2012). No luck with knowledge? On a dogma of epistemology. *Philosophy and Phenomenological Research*, DOI: 10.1111/j.1933-1592.2012.00622.

Becker, K. (2007). *Epistemology modalized*. Routledge.

Becker, K. (2008). Epistemic luck and the generality problem. *Philosophical Studies*, 139, 353-366.

Bernecker, S. (2008). Agent reliabilism and the problem of clairvoyance. *Philosophy and Phenomenological Research*, 76, 164-72.

Bogardus, T. (2012). Knowledge under threat. *Philosophy and Phenomenological Research*, DOI: 10.1111/j.1933-1592.2011.00564.x.

BonJour, L. (1980). Externalist theories of empirical knowledge. *Midwest Studies in Philosophy*, 5, 53-73.

BonJour, L. (1985). *The structure of empirical knowledge*. Harvard University Press.

Brendel, E. 2009. The epistemic function of virtuous dispositions. In Damschen, G., Schnepf, R. & Stueber, K. (Eds.): *Debating dispositions: Issues in metaphysics, epistemology and philosophy of mind*. DeGruyter.

Breyer, D. & Greco, J. (2008). Cognitive integration and the ownership of belief: Response to Bernecker. *Philosophy and Phenomenological Research*, 76, 173-184.

Breyer, D. (2010). Reflective luck and belief ownership. *Acta Analytica*, 25, 133-154.

Broncano-Berrocal, F. (2013). Lies and deception: A failed reconciliation. *Logos & Episteme. An International Journal of Epistemology*, 4, 227–230.

Broncano-Berrocal, F. (*forthcoming*). Is safety in danger? *Philosophia*.

Brueckner, A. & Buford, C. (2013). Becker on epistemic luck. *Philosophical Studies*, 163, 171-175.

Carlson, E. (2000). Incompatibilism and the transfer of power necessity. *Noûs*, 34, 277-290.

Carnap, R. (1950). *Logical foundations of probability*. The University of Chicago Press.

Carter, J. A. (2010). Anti-luck epistemology and safety's (recent) discontents. *Philosophia*, 38, 517-532.

Carter, J. A. (2013). A problem for Pritchard's anti-luck virtue epistemology. *Erkenntnis*, 78, 253-275.

Chisholm, R. M. (1977). *Theory of knowledge*. Prentice-Hall.

Clark, A. (1997). *Being there: Putting body, world, together again*. MIT Press.

Clark, A. E. & Lelkes, O. (2009). Let us pray: religious interactions in life satisfaction. *PSE Working Papers 2009-10*, PSE (Ecole normale supérieure).

Coffman, E. J. (2007). Thinking about luck. *Synthese*, 158, 385-398.

Coffman, E. J. (2009). Does luck exclude control? *Australasian Journal of Philosophy*, 87, 499-504.

Comesaña, J. (2005). Unsafe knowledge. *Synthese*, 146, 395-404.

Conee, E. & Feldman, R. (2004). *Evidentialism: Essays in epistemology*. Oxford University Press.

Craig, E. (1990). *Knowledge and the state of nature: An essay in conceptual synthesis*. Oxford University Press.

Dawkins, R. (1976). *The selfish gene*. Oxford University Press.

Dennett, D. (1984). *Elbow room: The varieties of free will worth wanting*. MIT Press.

DeRose, K. (1995). Solving the skeptical problem. *Philosophical Review*, 104, 1-52.

Dretske, F. (1981). *Knowledge and the flow of information*. MIT Press.

Eagleman, D. (2011). *Incognito: The secret lives of the brain*. Pantheon.

Égré, P. (2008). Reliability, margin for error and self-knowledge. In Hendricks, V. & Pritchard, D. (Eds.): *New waves in epistemology*. Palgrave Macmillan

Engel, M. (1992). Is epistemic luck compatible with knowledge? *Southern Journal of Philosophy*, 30, 59-75.

Fara, M. (2008). Dispositions. *Stanford Encyclopedia of Philosophy*.

Fischer, J. M. & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge University Press.

Fischer, J. M. & Ravizza, M. (2000). Précis of Responsbility and control: A theory of moral responsibility. *Philosophy and Phenomenological*, 61, 441-445.

Fischer, J. M. (2012). *Deep control: Essays on free will and value*. Oxford University Press.

Foley, R. (2004). A trial separation between the theory of knowledge and the theory of justified belief. In Greco, J. (Ed.): *Ernest Sosa and his critics*. Blackwell.

Fowler, H. N. (Trans.) (1921). *Plato: In twelve volumes, with an English translation, volume 12*. Harvard University Press.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, 23, 121-123.

Goldman, A. (1986). The cognitive and social sides of epistemology. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 2, 295-311.

Goldman, A. (2009). Williamson on knowledge and evidence. In Greenough, P. & Pritchard, D. (Eds.): *Williamson on knowledge*. Oxford University Press.

Goldman, A. I. (1967). A causal theory of knowing. *Journal of Philosophy*, 64, 357-372.

Goldman, A. I. (1975). Innate knowledge. In Stich, S. P. (Ed.): *Innate ideas*. University of California Press.

Goldman, A. I. (1976). Discrimination and perceptual knowledge. *Journal of Philosophy*, 73, 771-791.

Goldman, A. I. (1979). What is justified belief? In Pappas, G. (Ed.): *Justification and knowledge*. D. Reidel.

Goldman, A. I. (1999). Internalism exposed. *Journal of Philosophy*, 96, 271-293.

Goldman, A.I. (1986). *Epistemology and cognition*. Harvard University Press.

Greco, J. (1995). A second paradox concerning responsibility and luck. *Metaphilosophy*, 26, 81-96.

Greco, J. (2000). Putting skeptics in their place: *The nature of skeptical arguments and their role in philosophical inquiry*. Cambridge University Press.

Greco, J. (2003). Knowledge as credit for true belief. In DePaul, M. & Zagzebski, L. (Eds.): *Intellectual virtue: Perspectives from ethics and epistemology*. Oxford University Press.

Greco, J. (2004). How to preserve your virtue while losing your perspective. In Greco, J. (Ed.): *Ernest Sosa and his critics*. Blackwell.

Greco, J. (2010). *Achieving knowledge: A virtue-theoretic account of epistemic normativity*. Cambridge University Press.

Greco, J. (2012). A (different) virtue epistemology. *Philosophy and Phenomenological Research*, 85, 1-26.

Gundersen, L. (2003). *Dispositional theories of knowledge: A defence of aetiological foundationalism*. Ahsgate.

Gundersen, L. (2010). Tracking, epistemic dispositions and the conditional analysis. *Erkenntnis*, 72, 353-364.

Haddock, A. Millar, A. & Pritchard, D. (2010). *The nature and value of Knowledge: Three investigations*. Oxford University Press.

Harman, G. (1973). *Thought*. Princenton University Press.

Hawley, K. (2003). Success and knowledge-how. *American Philosophical Quarterly*, 40, 19-31.

Hawthorne, J. & Lasonen-Aarnio, M. (2009). Knowledge and objective chance. In Greenough, P. & Pritchard, D. (Eds.): *Williamson on knowledge*. Oxford University Press.

Hawthorne, J. (2004). *Knowledge and lotteries*. Oxford University Press.

Henderson, D. & Horgan, T. (2009). Epistemic virtues and cognitive dispositions. In Damschen, G., Schnepf, R. & Stueber, K. (Eds.): *Debating dispositions: Issues in metaphysics, epistemology and philosophy of mind*. DeGruyter.

Hendricks, V. F. (2006). *Mainstream and formal epistemology*. Cambridge University Press.

Hieronymi, P. (2006). Controlling attitudes. *Pacific Philosophical Quarterly*, 87, 45-74.

Hieronymi, P. (2008). Responsibility for believing. *Synthese*, 161, 357-373.

Hieronymi, P. (2009). Two kinds of agency. In O'Brien, L. & Soteriou, M. (Eds.). *Mental action*. Oxford University Press.

Hiller, A. & Neta, R. (2007). Safety and epistemic luck. *Synthese*, 158, 303-313.

Hopgood, A.A. (2012). *Intelligent systems for engineers and scientists*. CRC Press.

Johnston, M. (1992). How to speak of the colors. *Philosophical Studies*, 68, 221-263.

Kelp, C. (2009). Knowledge and safety. *Journal of Philosophical Research*, 34, 21-31.

Kelp, C. (2012). Knowledge: The safe-apt view. *Australasian Journal of Philosophy*, DOI: 10.1080/00048402.2012.673726

Kment, B. (2006). Counterfactuals and the analysis of necessity. *Philosophical Perspectives*, 20, 237-302.

Kolodny, N. (2008). Why be disposed to be coherent? *Ethics*, 118, 437-463

Kornblith, H. (2008). Knowledge needs no justification. In Smith, Q. (Ed.): *Epistemology: new essays*. Oxford University Press.

Kripke, S. (2011). *Philosophical troubles: Collected papers, volume I*. Oxford University Press.

Lackey, J. (2006). Pritchard's epistemic luck. *Philosophical Quarterly*, 56, 284-289.

Lackey, J. (2007). Why we don't deserve credit for everything we know. *Synthese*, 158, 345-361.

Lackey, J. (2008). What luck is not. *Australasian Journal of Philosophy*, 86, 255-267.

Lackey, J. (2009). Knowledge and credit. *Philosophical Studies*, 142, 27-42.

Lackey, J. (2013). Lies and deception: An unhappy divorce. *Analysis*, 73, 236-248.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, 32, 311-328.

Latus, A. (2003). Constitutive luck. *Metaphilosophy*, 34, 460-475.

Lehrer, K. & Cohen, S. (1983). Justification, truth, and coherence. *Synthese*, 55, 191-207.

Lehrer, K. (1990). *Theory of knowledge*. Westview Press.

Lepock, C. (2011). Unifying the intellectual virtues. *Philosophy and Phenomenological Research*, 83, 106-28.

Levy, N. (2009). What, and where, luck is: A response to Jennifer Lackey. *Australasian Journal of Philosophy*, 87, 489-497.

Levy, N. (2011). *Hard luck: How luck undermines free will and moral responsibility*. Oxford University Press.

Lewis, D. (1973). *Counterfactuals*. Blackwell.

Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy*, 74, 549-567.

Manley, D. (2007). Safety, content, apriority, self-knowledge. *Journal of Philosophy*, 104, 403-23.

Mellor, C. S. (1970). First rank symptoms of schizophrenia. *British Journal of Psychiatry*, 117, 15-23.

Mellor, D. H. (2005). *Probability: A philosophical introduction*. Routledge.

Meredith, A & Stein, B. E. (1993). *The merging of the senses*. MIT Press.

Montmarquet, J. (1993). *Epistemic virtue and doxastic responsibility*. Rowman & Littlefield.

Moore, A. W. (1990). A Kantian view of moral luck. *Philosophy*, 65, 297-321.

Murray, R. M. (Ed.) (2003). *Control in an information rich world: Report of the panel on future directions in control, dynamics and systems*. Society for Industrial and Applied Mathematics.

Nagel, T. (1979). *Mortal questions*. Cambridge University Press.

Neta, R. & Rohrbaugh, G. (2004). Luminosity and the safety of knowledge. *Pacific Philosophical Quarterly*, 85, 396-406.

Nozick, R. (1981). *Philosophical explanations*. Harvard University Press.

Olsson, E. J. (2010). *Coherentism*. In Bernecker, S. & Pritchard, D. (Eds): The Routledge companion to epistemology. Routledge.

Owens, D. (1992). *Causes and coincidences*. Cambridge University Press.

Peacocke, C. (1999). *Being known*. Oxford University Press.

Plantinga, A. (1993a). *Warrant and proper function*. Oxford University Press.

Plantinga, A. (1993b). *Warrant: The current debate*. Oxford University Press.

Pritchard, D. (2003). Virtue epistemology and epistemic luck. *Metaphilosophy*, 34, 106-130.

Pritchard, D. (2005). *Epistemic luck*. Oxford University Press.

Pritchard, D. (2007). Anti-luck epistemology. *Synthese*, 158, 277-297.

Pritchard, D. (2008). Greco on knowledge: Virtues, contexts, achievements, *The Philosophical Quarterly*, 58, 437-47.

Pritchard, D. (2009a). Safety-based epistemology: Whither now? *Journal of Philosophical Research*, 34, 33-45.

Pritchard, D. (2009b). Apt performance and epistemic value. *Philosophical Studies*, 143, 407-416.

Pritchard, D. (2010a). Achievements, luck and value. *Think*, 25, 1-12.

Pritchard, D. (2010b). Cognitive ability and the extended cognition thesis. *Synthese*, 175, 133-151.

Pritchard, D. (2012a). Wittgenstein and the groundlessness of our believing. *Synthese*, 189, 255-272.

Pritchard, D. (2012b). Anti-luck virtue epistemology. *The Journal of Philosophy*, 109, 247-279.

Pritchard, D. (2012c). In defence of modest anti-luck epistemology. In Black, T. & Becker, K (Eds.): *The sensitivity principle in epistemology*. Cambridge University Press.

Pryor, J. (2004). Comments on Sosa's "Relevant alternatives, contextualism included". *Philosophical Studies*, 119, 67-72.

Rabinowitz, D. (2011). The safety condition for knowledge. *Internet Encyclopedia of Philosophy*.

Raz, J. (2011). Being in the world. In De Gaynesford, M. (Ed.): *Agents and their actions*. Wiley-Blackwell.

Rescher, N. (1995). *Luck: The brilliant randomness of everyday life*. Farrar Strauss Giroux.

Riggs, W. (2002). Reliability and the value of knowledge. *Philosophy and Phenomenological Research*, 64, 79-96.

Riggs, W. (2007). Why epistemologists are so down on their luck. *Synthese*, 158, 329-344.

Riggs, W. (2009). Knowledge, luck, and control. In Haddock, A., Millar, A. & Pritchard, D. (Eds.): *Epistemic Value*. Oxford University Press.

Roberts, T. (2012). Taking responsibility for cognitive extension. *Philosophical Psychology*, 25, 1-11.

Rosenkranz, S. (2007). Agnosticism as a third stance. *Mind*, 116, 55-104.

Roush, S. (2005). *Tracking truth: Knowledge, evidence, and science*. Oxford University Press.

Rowlands, M. (2009). Extended cognition and the mark of the cognitive, *Philosophical Psychology*, 22, 1-19.

Russell, B. (1948). *Human knowledge*: *Its scope and its limits*. Allen & Unwin.

Sainsbury, M. (1997). Easy possibilities. *Philosophy and Phenomenological Research*, 57, 907-919.

Sarno, S., Erasmus, L., Lipp, B., & Schlaegel, W. (2003). Multisensory integration after traumatic brain injury: a reaction time study between pairings of vision, touch and audition. *Brain Injury*, 17, 413-426.

Shope, R. K. (1983). *The analysis of knowledge: A decade of research*. Princeton University Press.

Sosa, E. (1980). The raft and the pyramid: Coherence versus foundations in the theory of knowledge. *Midwest Studies in Philosophy*, 5, 3-26.

Sosa, E. (1991). *Knowledge in perspective: Selected essays in epistemology*. Cambridge University Press.

Sosa, E. (1994). Virtue perspectivism: A response to Foley and Fumerton. *Philosophical Issues*, 5, 29-50.

Sosa, E. (1999). How to defeat opposition to Moore. *Philosophical Perspectives*, 13, 137-49.

Sosa, E. (2002). Tracking, competence and knowledge. In Moser, P. (Ed.): *The Oxford handbook of epistemology*. Oxford University Press.

Sosa, E. (2003). Beyond internal foundations to external virtues. In BonJour, L. & Sosa, E. *Epistemic justification: Internalism vs. externalism, foundations vs. virtues*. Blackwell.

Sosa, E. (2004). Replies. In Greco, J. (Ed.): *Ernest Sosa and his critics*. Blackwell.

Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge, volume I*. Oxford University Press.

Sosa, E. (2009). *Reflective knowledge: Apt belief and reflective knowledge, volume II*. Oxford University Press.

Sosa, E. (2010). How competence matters in epistemology. *Philosophical Perspectives*, 24, 465-475.

Sosa, E. (2011). *Knowing full well*. Princenton University Press.

Statman, D. (1991). Moral and epistemic luck. *Ratio*, 4, 146–156.

Steglich-Petersen, A. (2010). Luck as an epistemic notion. *Synthese*, 176, 361-377.

Strzałko, J., Grabski, J., Perlikowski, P., Stefański, A. & Kapitaniak, T. (2008). The dynamics of coin tossing is predictable. *Physics Reports*, 469, 1-34.

Turri, J. (2011). Manifest failure: The Gettier problem solved. *Philosophers' Imprint*, 11, 1-11.

Turri, J. (2013). Bi-level virtue epistemology. In Turri, J. (Ed.): *Virtuous thoughts: The philosophy of Ernest Sosa*. Philosophical Studies Series, 119, 147-164.

Turri, J. (forthcoming). Knowledge as achievement, more or less. In Fernández, M. A. (Ed.): *The present and future of virtue epistemology*. Publisher TBD.

Unger, P. (1968). An analysis of factual knowledge. *Journal of Philosophy*, 65, 157-170.

Vaesen, K. (2011). Knowledge without credit, exhibit 4: extended cognition. *Synthese*, 181, 515-529.

Vogel, J. (2007). Subjunctivitis. *Philosophical Studies*, 134, 73-88.

Warfield, T. (2005). Knowledge from falsehood. *Philosophical Perspectives*, 19, 405-416.

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

Williamson, T. (2009). Probability and danger. *The Amherst Lecture in Philosophy*, 4, 1-35.

Williamson, T. (2009). Replies to critics. Williamson on knowledge and evidence. In Greenough, P. & Pritchard, D. (Eds.): *Williamson on knowledge*. Oxford University Press.

Zagzebski, L. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge University Press.

Zagzebski, L. (2003). The search for the source of epistemic good. *Metaphilosophy*, 34, 12-28.

Zimmerman, M. (1987). Luck and moral responsibility. *Ethics*, 97, 374-386.

COLOPHON

This thesis was typeset in LyX, available from www.lyx.org. using the typographical style `classicthesis` developed by André Miede, available for both LaTeX and LyX from http://code.google.com/p/classicthesis/.