UNIVERSITAT DE
BARCELONA

Domain-general perspectives on the neurocognitive specialization of language

Gonzalo Castillo

DOMAIN-GENERAL PERSPECTIVES ON THE NEUROCOGNITIVE SPECIALIZATION OF LANGUAGE

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Linguistics
Cognitive Science and Language
Department of General Linguistics
Universitat de Barcelona

GONZALO CASTILLO
SEPTEMBER 2017

ADVISOR:
DR CEDRIC BOECKX

Abstract

This dissertation proposes that the frontotemporal network involved in the neural substrate of language processing is as a subset of a domain-general system mediated by predictive processing mechanisms. Instead of attributing independent functions to each of its patterns of activity, it redescribes them as deriving from differences in the performance of this system when attempting integrations. In particular, it proposes that dorsal-ventral interactions are behind all linguistic and cognitive functions, representing as a single stream the exchange of feedforward (prediction error) and feedback (prediction) information under differing degrees of estimated certainty. This exchange is in turn manifested across a rostrocaudal hierarchy ending in pars triangularis, which connects the current processing context with retrieved task sets for its more efficient, proactive processing, and is generally involved in cognitive control.

The proposed domain-general system produces different cross-network interactivity patterns based on the ongoing levels of expected and unexpected uncertainty, sustaining language and cognition. Evidence for these is gathered from event-related potentials, oscillatory activity, neurochemistry, network neuroscience, and a broad range of cognitive neuroscience studies.

The last part of the dissertation studies the related problem of how language and speech could evolve simultaneously as a result of continuous, quantitative changes taking place in the aforementioned domain-general system. This section proposes a new framework to study the evolution of vocal learning abilities in connection with domain-general cognition, based on primate neural scaling rules and connectivity enhancements, and argues against comparative perspectives that only focus on animal vocalizations in order to explain human specificity.

Contents

List of figures

***

List of abbreviations

a: anterior

ACC: anterior cingulate cortex

ACh: acetylcholine

AE schemata: actor-event schemata

AF: arcuate fasciculus

AP: anterior positivity

ATL: anterior temporal lobe

BA: Brodmann area

BOLD: blood oxygenation level dependent signal

CON: cingulo-opercular network

d: dorsal

dm: dorsomedial

DMN: default mode network

E (Type): expressive

ECFS: extreme capsule fiber system

ECoG: electrocorticography

EEG: electroencephalogram

ERP: event-related potential

EU: expected uncertainty

fMRI: functional magnetic resonance imaging

FOP: frontal operculum

FPN: frontoparietal network

FTN: frontotemporal network

HVC: HVC

IFG: inferior frontal gyrus

IPL: inferior parietal lobe

l: lateral

L (Type): lexical

LATL: left anterior temporal lobe

LIFS: left inferior frontal sulcus

LMLF: left middle longitudinal fasciculus

m: medial

MAN: magnocellular nucleus of anterior nidopallium

ms: milliseconds

MTG: middle temporal gyrus

MTL: medial temporal lobe

N: negative

NC-state: network connectivity state

NE: norepinephrine

NLc: lateral neostriatum

NP: noun phrase

OFC: orbitofrontal cortex

p: posterior

P: positive

PFC: prefrontal cortex

PCC: posterior cingulate cortex

RA: nucleus robustus archistrialis

RATL: right anterior temporal lobe

ROI: region of interest

SFG: superior frontal gyrus

SOA: stimulus onset asynchrony

STS: superior temporal sulcus

SVO: subject-verb-object

UF: uncinate fasciculus

UU: unexpected uncertainty

v: ventral

vm: ventromedial

VP: verb phrase

Introduction

Neurolinguistics often defines the brain correlates of language based on which regions closely track linguistic stimuli during comprehension tasks. This view has confined language to a frontotemporal network, which has been defined as domain-specific or specialized for language, and even as a human-specific adaptation that is not shared with other primates. This dissertation is based on the alternative point of view: that activations that follow stimuli do not represent everything that supports a specific behavior at the brain level, and that the capabilities of specialized neural substrates are inherited from the domain-general cognitive architecture that hosts them, which is responsible for their development, and does not stop playing a necessary role in sustaining their functions.

The big question asked in this dissertation is the following: how does the frontotemporal "language network" interact with the domain-general architecture of the brain? The general answer that is provided is that active predictive coding can account for the observed patterns of cross-network interactivity without resorting to domain-specific terminology, which has the advantage of connecting neurolinguistics with broader evolutionary and developmental questions, and with insights from other domains of inquiry. In this framework, the underlying assumption is that the cognitive system constantly tries to learn models of linguistic input to actively predict or simulate incoming signals, and that the efficiency of the match between model and bottom-up input determines activation patterns.

For the purpose of tractability, this question is divided into smaller ones, which occupy each of the three chapters that compose it. Chapter 1 focuses on investigating the relationship between the frontotemporal and frontoparietal networks, specifically on how synchronizations between both networks can be characterized as the contextual recruitment of representation and error units under conditions of increased uncertainty. In doing so, it questions theories that consider that these networks are performing fundamentally different functions, like syntax and working memory.

Chapter 1 also proposes that the functions attributed to the dorsal stream of the frontotemporal network depend on the activity of the ventral stream, which is defined as a higher-level interface between the domain-specific, stimulus-tracking dorsal stream, and the domain-general cognitive architecture encompassing uncorrelated cortical and subcortical activity. Evidence for this single-stream framework is offered from a neuroconstructivist perspective (Karmiloff-Smith, 2015), which suggests an initially more combined developmental trajectory for both networks, and also by providing a comprehensive analysis of how language-related evoked potentials, spontaneous oscillatory activity, and the neurochemistry behind uncertainty can support it.

Chapter 2 focuses on characterizing the role of the ventral stream further by asking three questions: is there an actual brain basis for the separation between local and complex syntax? What is the function of increased ventral activity in the absence of dorsal engagement? And: what is the relationship between the ventral stream and domain-general networks like the cingulo-opercular and default-mode networks? The hypothesis that is proposed is that the estimation of certainty about predictions is behind all the different patterns of activity that motivate these distinctions.

Thus, the local/complex distinction and its association with different streams is questioned on the basis that dorsal-ventral interactions seem to be behind all linguistic operations, and activation patterns like ventral as opposed to dorsal activity emerge as performance differences caused by the specifics of tasks. Also, the functional connectivity of complex syntactic operations is analyzed from the point of view of domain-general functional networks, showing how this label (and its associated brain correlates) cannot account for all the domain-general data, while the current framework fares better.

Chapter 3 moves on to evolutionary questions. The starting hypothesis is that the adoption of a domain-general perspective on the trait of vocal learning (a focus on its phylogenetically shared basis) can reveal more relevant cross-species comparisons than its understanding as a species-specific adaptation. The first part of the chapter is a critical analysis of the shortcomings of domain-specific hypotheses on the origins of language based on vocal learning. The second part offers a domain-general alternative. The main proposal is that enhancements in domain-general cognition necessarily affect modular behaviors by making them less modular or shielding them against further modularization. This has similar consequences for the domain-specific brain systems that form the core of these behaviors, as they become more interconnected, and able to participate in a more widespread range of functions.

This approach allows the proposal of a new kind of vocal learning continuum that can capture the different manifestations of vocal imitation across species, suggests that the frontoparietal-frontotemporal relationship studied in Chapter 1 can be directly compared with the parrot brain, and leads to a hypothesis for how vocal learning and language could have evolved simultaneously as part of the same set of changes, as opposed to at different stages.

In sum, the original contributions offered by this dissertation are a reinterpretation of two-stream parsing models of the frontotemporal network on the basis of a single-stream model based on the tenets of hierarchical predictive coding (Chapters 1 and 2), and a hypothesis on how language and vocal learning could have evolved simultaneously as part of evolutionary changes affecting the quantitative scale of the former system (Chapter 3). Together, these two ideas aim to establish more interdisciplinary bridges between neurolinguistics and other domains of cognitive neuroscience.

***

1. The dorsal stream and the frontoparietal network – a domain-general, hierarchical predictive coding perspective

The main goal of this chapter is the introduction of a new framework to study the functional relationship that exists between two functional networks that are widely thought to be involved in language processing, the frontoparietal and frontotemporal networks. The specific function of each network, their necessity for language, and the characterization of how they relate to each other during online processing and ontogenetically are currently a matter of debate in the neurolinguistics community. Finding answers to these questions depends on building a model that characterizes language at multiple levels of analysis and from an interdisciplinary perspective. The contribution to this quest made by this chapter is showing how the problem can be studied from a domain-general perspective, that is, from the point of view of how the brain supports the performance of any cognitive task, as opposed to specifically linguistic tasks. This point of view has the advantage that insights obtained in other areas of cognitive neuroscience can inform our conclusions about language. The main one discussed in this chapter is that if the brain approximates a Bayesian inferential machine (Jaynes, 1988), the relationship between the frontoparietal and frontotemporal networks can also be understood in this way.

This chapter is structured as follows. It starts by introducing the reader to what we know about the frontoparietal-frontotemporal relationship so far, summarizing it into three different hypotheses: the working memory hypothesis, the cognitive control hypothesis, and the hierarchical predictive coding hypothesis, defended here. After discussing the developmental motivations for choosing a hierarchical predictive coding account over the other two, the following section examines through the predictive coding lens a fundamental question arising from the neurolinguistics literature, namely the processing of non-adjacent relationships. Next, the topic of how event-related potentials may support a decomposition of predictive mechanisms into more fundamental components is addressed. This allows us to establish a direct connection between prediction and the relationship between both networks, which supports the existence of a continuum based on how much ventral predictions are constrained by dorsal activity.

1.1. Two functional networks involved in language – three hypotheses to explain their relationship

Normally called the *language system*, the frontotemporal network (FTN) is a mostly left-lateralized functional network that includes the temporal lobe (posterior, middle, and anterior regions), inferior, middle and orbital frontal gyri, and the angular gyrus, as well as parts of parietal cortex, dorsal striatum and right cerebellum (Fedorenko et al., 2010, 2011, 2012; Fedorenko & Thompson-Schill, 2014; Fedorenko, 2014). The FTN is organized into various white matter tracts that can be grouped into two main streams: a dorsal stream that connects the inferior frontal gyrus (BA 44) to posterior temporal cortex, traversing premotor cortex, inferior parietal lobule, and superior temporal cortex; and a ventral stream that connects inferior frontal gyrus (BA 45) with anterior superior temporal cortex (Bornkessel-Schlesewsky et al., 2015). These streams are the linguistic variants of a domain-general architectural property, as this distinction also exists for the domains of vision and audition (Mishkin & Ungerleider, 1982; Goodale & Milner, 1992; Bornkessel-Schlesewsky & Schlesewsky, 2013).

The domain-specificity of the FTN is attested by the fact that its regions selectively activate for and closely track linguistic input. This is especially important within inferior frontal gyrus, where language-selective regions belonging to the FTN are adjacent to domain-general regions belonging to the frontoparietal network (Fedorenko et al., 2012).

Functionally, the task of mapping individual concepts of theoretical linguistics onto the FTN is problematic (cf. Poeppel, 2012; Blank et al., 2016; Strijkers & Costa, 2016), but the consensus seems to be that it is at least possible to distinguish regions that are more associated with meaning (temporal pole, anterior superior temporal sulcus, temporo-parietal junction), sound (superior temporal gyrus), and their combination (inferior frontal and posterior temporal cortex) (Pallier et al., 2011; Friederici, 2012; Goucha & Friederici, 2015). Some researchers advocate for a less language-specific functional division based on the domain-general computational properties of the two-stream model. Thus, the dorsal stream of the FTN would be in charge of order-sensitive computations, including auditory-to-motor transformations (Hickok & Poeppel, 2007; Saur et al., 2008; Bornkessel-Schlesewsky & Schlesewsky, 2013; Bornkessel-Schlesewsky et al., 2015), while the ventral stream would handle more abstract, order-insensitive computations (Rauschecker, 2012; Bornkessel-Schlesewsky et al., 2015). In this model, the role of prefrontal cortex is not syntactic combination, but cross-stream integration (i.e., it would be an interface between the dorsal and ventral streams).

The frontoparietal network (FPN) or *multiple-demand system* (Duncan, 2010, 2013) is a functional network comprising the frontal gyri, inferior and superior parietal lobes, supplementary and presupplementary motor areas, precentral gyri, as well as parts of the basal ganglia, thalamus and cerebellum (Blank et al., 2014; Duncan, 2010, 2013; Fedorenko et al. 2013). The FPN is the most prominent domain-general network, as its increased activity is involved in multiple functions like cognitive control within all sensory domains (Vincent et al., 2008), conscious awareness (Dehaene & Naccache, 2001; Dehaene & Changeux, 2011), top-down attentional processes (Corbetta & Shulman, 2002; Duncan, 2013), the hierarchical ordering across time of plans and actions (Koechlin et al., 2003; Duncan, 2010), and language (Fedorenko et al., 2012; 2013; Fedorenko, 2014). It is also associated with the notion of fluid intelligence (Woolgar et al., 2010).

The FPN is proposed to be structured hierarchically, so that motor acts encoded posteriorly would progressively turn into more abstract and temporally distant goals in anterior regions (Koechlin & Summerfield, 2007; Badre & D'Esposito, 2009; Wacongne et al., 2011). This hierarchy entails that the FPN is especially sensitive to conditions of processing difficulty or effort, which tend to be associated with more anterior activity (Duncan, 2010). For example, in the domain of vision, Woolgar et al. (2011) demonstrate that by introducing noise in a rule-based stimulus response task, activity within the domain-specific visual system (occipital lobe) decreases as activity within prefrontal cortex (FPN) increases. Moreover, the activity of prefrontal cortex was shown to reflect the feature of the task that became problematic because of the noise (i.e., position), a phenomenon that is interpreted as an enhancement of representational capacity. This crucial property of the FPN is known as *dynamic coding* (Stokes et al., 2013, Pezzulo et al., 2015), a name that reflects the fact that frontoparietal neurons can flexibly adapt, via short-term plasticity or oscillatory mechanisms, to code for context-dependent rules and stimuli across multiple domains.

For language processing, both the FTN and FPN engage consistently and in tandem during language production. In language comprehension, however, the patterns of activity are more complex: while the FTN seems to closely track linguistic stimuli, the FPN displays inconsistent, uncorrelated patterns of activity that can change across individuals and trials (Blank et al., 2014; Blank & Fedorenko, 2016). Consistent with the functionality just ascribed to this network, the apparent cause of FPN engagement during language comprehension is the presence of linguistic stimuli that the subject tends to find more difficult to process. Fedorenko (2014) provides a list of such conditions: non-local syntactic dependencies (especially for aging individuals), ambiguity, pronouns with difficulties in reference attribution, sentences with grammatical errors, and the presence of noise or any suboptimal conditions for comprehension. Additionally, the processing of a second language is also associated with enhanced FPN activity (Jeon & Friederici, 2013, 2015).

This chapter considers three hypotheses that attempt to explain the FPN-FTN relationship:

The *working memory* hypothesis claims that frontal regions belonging to the FPN act as a language-specific storage buffer that keeps a syntactic object active while the rest of a sentence is parsed,[1] so that it can be recovered later when it is needed to establish a syntactic relationship (Gibson, 1998; Caplan & Waters, 1999; Fiebach et al., 2001, 2002, 2005; Santi & Grodzinsky, 2007; Makuuchi et al., 2009; Makuuchi & Friederici, 2013). This hypothesis assumes that the FPN and FTN have a very different computational nature, as the former would be involved in short-term storage, and the latter in the processing of hierarchical dependencies (syntax).[2]

The *cognitive control* hypothesis understands the FPN as an optional top-down controller of domain-specific systems like the FTN, which normally handle the actual processing in a bottom-up manner. Following the notion of dynamic coding, the assumption is that the FPN activates to sharpen the neural representations of the FTN when automatized processing within it is not enough to accurately represent the input. In the absence of these top-down effects, language comprehension is therefore understood as a mostly passive, automatic process based on the sound-guided reactivation of consolidated structures contained within the FTN (Fedorenko, 2014). Contrarily to the working memory hypothesis, the cognitive control hypothesis discards the existence of

---

[1] More specifically, some accounts argue that the working memory buffer only exists for movement or long-distance (intrasentential) dependencies (e.g., Fiebach et al., 2005), while others claim a more generic syntactic mechanism that is also active for any hierarchical relationship between elements (Caplan & Waters, 1999). The present dissertation, on the other hand, understands this difference as a quantitative one, relying on the same neural mechanisms (namely ontogenetically specialized dorsal-ventral integrations).

[2] It should be noted that current versions of this hypothesis have to account for three important discoveries: that all regions of the dorsal stream are sensitive to complexity increases (Blank et al., 2016), that prefrontal regions previously assumed to be unitary show a very intricate subdivision (Amunts & Zilles, 2012), and that prefrontal regions that are sensitive to linguistic input can be divided into language-selective and domain-general (Fedorenko et al., 2012). This is why, if the reader visits older papers, the working memory hypothesis is presented as involving Broca's area only, assumed to possess a unitary anatomy and functionality, and to belong to the FTN only. The hypothesis discussed here, on the other hand, is that domain-general regions of Broca's area handle working memory in opposition to domain-specific regions handling hierarchy.

language-specific structures and mechanisms within the FPN (but not necessarily within the FTN). In both hypotheses, the FPN plays an ancillary role for language.

Lastly, the *hierarchical predictive coding* hypothesis, defended here, argues that the FPN-FTN relationship is essential for all instances of language processing, including production and comprehension (e.g. McClelland et al., 2006), as both networks form part of a predictive coding hierarchy in which the activity of all networks matters. This chapter also claims that the FTN is a domain-specific specialization that arises ontogenetically as part of a process of neural reuse (Anderson, 2010, 2016; Christiansen & Müller, 2015), and therefore inherits the same basic computational operation of reducing prediction error. In contrast with the other two hypotheses, this account defends the existence of a single function and a continuous relationship that does not fundamentally change regardless of the specific linguistic context (comprehension and production)[3], or the activation patterns observed across each network, which only reflect performance differences.

The next section explores the notion of hierarchical predictive coding in more detail, and explains the developmental motivations for preferring this account over the other two.

---

[3] This is in agreement with other unified accounts of production and comprehension (e.g. Pickering & Garrod, 2007, 2013; MacDonald, 2013), and with predictive coding reinterpretations of the mirror neuron literature, where action understanding and production are interwoven, but mirror neurons, like the FTN's dorsal stream, are just the lower-level, stimulus-tracking subset of the whole hierarchy (Kilner et al., 2007; Kilner & Frith, 2008).

1.2. Hierarchical predictive coding and the developmental relationship between domain-general and domain-specific systems

Hierarchical predictive coding is a model of how the brain supports cognition, perception, attention and action (Mumford, 1992; Rao & Ballard, 1999; Friston, 2005, 2010; Friston & Kiebel, 2009; Friston et al., 2010). It rests on the claim that the brain works by constantly making probabilistic, knowledge-driven inferences on raw sensory data, developing top-down *predictions* that are actively contrasted with it, and subsequently revised or discarded as more evidence accumulates (Helmholtz, 1860/1962; Clark, 2013).

Hierarchical predictive coding is proposed to be implemented neurally by a cortex that is structured hierarchically, and where information flows bidirectionally, through *feedforward* and *feedback* connections (Barbas, 2015). This hierarchy progressively summarizes and abstracts the high complexity and noise of sensory input, with each level generating a model of how the level below will behave. The reception of this feedback suppresses (“explains away”, Rao & Ballard, 1999) the bottom-up input that can be predicted by using it. Only what cannot be predicted, known as surprisal or prediction error, is sent to the higher level. Thus, prediction error is the only information that gets transmitted by feedforward connections, and only to the extent that feedback connections are not able to suppress it. Instead of maintaining multiple versions of the same information at different levels, activation decreases as the higher-levels find states that steadily inhibit the lower-levels.

In current implementations, each level of the hierarchy is thought to possess two kinds of units of computation (neurons): prediction and error units. *Error units* compute prediction error, and transmit it to the level above. *Prediction units* encode predictions at their own level, and suppress lower-level error units when predictions are effective. Supragranular pyramidal neurons are thought to play the role of error units, while infragranular pyramidal neurons would play the role of representation units (Bastos et al., 2012).

Models that consider cognition as being driven by prediction have become popular in cognitive neuroscience, and language is not the exception (e.g., Townsend & Bever, 2001; Bever & Poeppel, 2010; Bornkessel-Schlesewsky & Schlesewsky, 2013; Hickok, 2012; Lupyan & Clark, 2015). Multiple experiments show that the expectations derived from the context of an utterance can pre-activate linguistic representations, affecting reaction times, activation patterns and event-related potentials during language comprehension (Federmeier, 2007; Kutas & Federmeier, 2011; Van Petten & Luca, 2012; Kuperberg & Jaeger, 2016; Dröge et al., 2016; Zarcone et al., 2016). However, it should be noted that the hypothesis of hierarchical predictive coding goes beyond the proposal that the brain can anticipate some stimuli provided the existence of a sufficiently rich context. Hierarchical predictive coding is instead an explanation of how neural activity as a whole works, including when processing unexpected stimuli, which should entail a larger prediction error, but be solved in the same way as any other input. It is also a model of how learning works, as recurrent feedback connections can learn, store, and predict sequences of neural activity (Rao & Sejnowski, 2002), altering the weight given to certain patterns of prediction error and biasing the search for predictions. In short, the concept of neural prediction should not be confused with the concept of cognitive prediction, or *pre-activation*, even if the latter are also formed by neural predictive mechanisms.

It is in this context that the hypothesis that the FPN-FTN relationship is mediated by predictive coding is proposed. More specifically, the claim being made is that the different patterns of BOLD responses observed across both networks reflect what would be expected from the spreading of prediction error across a predictive coding hierarchy. Instead of standing for a new function, the engagement of the FPN would represent: a) a large surge of prediction error that recruits more prediction units referring to the combination of meaning and sound, or b) a large surge of prediction error that recruits more prediction units referring to processes that are parallel to the linguistic signal. While in the first case the FPN would seem to track linguistic stimuli like the FTN does, in the second case it would not (but the predictions made should still have an effect on the representation of the linguistic signal).

More specifically, the relationship that is proposed is one in which some regions of the FPN share the same cortical level as dorsal regions belonging to the FTN (specifically, the adjacent frontal and parietal regions; Fedorenko et al., 2012), whereas others occupy higher levels within the hierarchy (e.g., frontopolar cortex). Additionally, some regions of the FTN occupy lower levels than the lowest level of the FPN (e.g., superior temporal gyrus), and the ventral stream of the FTN occupies a higher level than the FPN. Whenever complexity increases enough to exceed domain-specific predictive capacities, prediction error will spread to adjacent FPN regions, and be contested via lateral connections. It will also spread upwards, and be contested by a combination of the activity of higher FPN and FTN levels via feedback connections. In sum, the FPN holds a continuous functional relationship with the FTN, instead of acting as a working memory buffer, or as the only source of feedback over bottom-up input.

The main motivation for this hypothesis is that the developmental trajectory of the left-dominant dorsal stream of the FTN is one characterized by increasing functional specialization for language processing departing from more distributed and metabolically costly domain-general activation patterns (Perani et al., 2011, Jeon & Friederici, 2013; Weiss-Croft & Baldeweg, 2015). This can be interpreted as indicating that domain-specific systems share fundamental computational properties with domain-general systems, acting as a shortcut for realizing them faster and more efficiently at the expense of adaptiveness and flexibility (Bullmore & Sporns, 2012; Frost et al., 2015).

The infant brain shows from birth lateralization patterns and an already present arcuate fascicle, which has been taken as supporting evidence for the existence of a “unique neural machinery” that supports language processing (Dehaene-Lambertz & Spelke, 2015). An alternative, domain-general view that acknowledges these data is to consider that these regions start as *language-sensitive* as opposed to *language-selective* (Johnson, 2011; Karmiloff-Smith, 2015), responding to language processing together with a variety of functions, in a way that is more characteristic of the domain-general FPN. The increased sensitivity of these regions over others may be due to lower-level causes such as the spectro-temporal properties of the linguistic signal, which shifts the balance of activity towards the left hemisphere and a postero-dorsal functional connection with posterior temporal cortex, and higher-level ones like the need to associate this signal with meaning, which promotes an antero-ventral functional connection with anterior and medial temporal lobe.

The process of language acquisition is one in which some language-sensitive regions become language-selective, and others become more domain-general as a result of losing the race to predict linguistic input in the most efficient way possible. The mechanisms behind this process should be the usual mechanisms of developmental change: local ones like cell-cell signaling, long-term potentiation and neural recycling, and global ones that affect long-distance connectivity like pruning, myelination and neural reuse (Dehaene & Cohen, 2007; Barrett, 2012; Anderson, 2010, 2016; Christiansen & Müller, 2015). In this way, a very widespread functional profile that involves bilateral regions belonging to the adult frontoparietal and default mode networks (Sheppard et al., 2012; Weiss-Croft & Baldeweg, 2015) turns into a more constrained and robust FTN that is especially sensitive to the language domain, and in the case of some subregions, answers to it exclusively (Fedorenko et al., 2012). Meanwhile, other regions that were initially language-sensitive do not progress further, freeing up resources for further specialization in other domains, or to dynamically code for prediction error and enrich predictions further (FitzGerald et al., 2014).

Thus, the process of specialization entails an important trade-off between speed and metabolic efficiency vs dynamic coding and flexibility. The functional associations that specialize should be the ones that deal with stimuli that are frequent and invariant enough for the reduction of flexibility to be worth it. Moreover, specialization should only be possible to the extent that enough domain-general functional associations remain, so that other functions that are less consolidated are not compromised. This implies that any deviation from canonical prediction error should trigger a cascade of activation that spreads towards more flexible neural assemblies, progressively situated within what we identify as the domain-general FPN.

In sum, this developmental framework suggests that the computational properties of specialized regions like the dorsal stream are inherited from the domain-general regions they specialize from, crucially including the reduction of prediction error. This feature requires that specialized processing cannot be understood without a necessary existence of domain-general contributions. In fact, in a constantly variable environment, the more complex a specific behavior becomes, the more important domain-general contributions may become, as more potential for unexpected variations would exist. In the following sections, these insights will be applied to explain the FPN-dorsal stream relationship observed across different studies, and how it relates to the ventral stream.

The next section offers a predictive coding perspective on one of the most studied phenomena in language, the processing of non-adjacent relationships, by paying special attention to how the FPN could relate to both the dorsal and ventral streams.

1.3. The predictive processing of non-adjacent relationships

This section offers from a predictive coding perspective a redescription of three known patterns of activity involved in language processing: dorsal stream engagement within the FTN, generally involved in language processing; modulations in dorsal stream engagement specifically associated with the processing of long-distance dependencies; and bilateral synchronization between the FTN and FPN associated with the processing of center-embeddings and, more contextually, with the processing of long-distance dependencies. The goal is to claim that 1) these patterns depend on a continuum of prediction error, reflecting performance differences within a single function, and 2) that the state of the ventral stream of the FTN needs to be taken into account to understand the emergence of one pattern over another.

The dorsal stream of the FTN, and particularly a language-specific subregion within BA 44 in inferior frontal gyrus (IFG) (Zaccarella & Friederici, 2015), is known to generally engage for phrase structure building. The working memory hypothesis claims that the operation of BA 44 is subserved by either language-specific or domain-general working memory buffers that aid its functioning. An example of the conditions in which these buffers are supposedly used is the processing of object-first relative sentences (e.g., "This is the boy who the girl kissed"). They are characterized by the need to associate a verb ("kissed") with an antecedent ("the boy") that is separated by intervening elements ("the girl"). This relative sentence is less frequent than the canonical subject-first relative sentence (e.g., "This is the boy who kissed the girl"), in which hierarchical relationships are established without having to "move" across intervening elements. Object-first relatives a processed more effortfully and slowly than subject-first relatives, a fact that has been attributed to multiple possible factors: the need to inhibit the canonical subject-first interpretation (e.g., Bahlmann et al., 2007), the fact that they may be less consolidated than subject-first sentences (e.g., Wells et al., 2009), or, crucially, the need to maintain the antecedent in working memory while processing the intervening materials so that a hierarchical relationship can be established (e.g. Gibson, 1998; Grodner & Gibson, 2005).

Neural evidence for the latter interpretation is found by Makuuchi and colleagues (Makuuchi et al., 2009; Makuuchi & Friederici, 2013; Santi et al., 2015), a series of results that help establish a functional division between the FTN and the FPN. Makuuchi et al.

(2009) test the following conditions, characterized by independently varying movement (distance) and hierarchical depth in center-embedded structures:

a) Hierarchy & long distance
[Maria, [die Hans, [der gut aussah], liebte], Johann geküsst hatte]
Maria who loved Hans who was good looking kissed Johann.

b) Hierarchy & short distance
[Maria, [die weinte], Johann geküsst hatte] und zwar gestern abend]
Maria who cried kissed Johann and that was yesterday night.

c) Linear & long distance
[Achim den großen Mann gestern am späten Abend gesehen hatte]
Achim saw the tall man yesterday late at night.

d) Linear & short distance
[Achim den großen Mann gesehen hatte und zwar am Abend]
Achim saw the tall man at night and that was late.

They observe that increases in distance (conditions *a* and *c*) are preferentially associated with activations in the dorsal bank of BA 45, a more anterior region than BA 44 within IFG, and LIFS, a region adjacent to IFG that can be characterized as belonging to the FPN. Meanwhile, increases in hierarchical complexity (conditions *a* and *b*) are reflected by more activity in BA 44. Also, when distance and hierarchical complexity increase together (condition *a*), activity in BA 44 and LIFS synchronizes. Makuuchi & Friederici (2013) further demonstrate that when processing object-first relatives there is an increased causal influence from LIFS to the IFG. The conclusions attained are that BA 44 processes hierarchical relationships, being sensitive to increases in hierarchical complexity, while LIFS acts as a working memory buffer that engages contextually when working memory is needed. This sets apart the FTN and the FPN as networks that are fundamentally different.

There is a number of results, coming from the same research group and others, that offer a more nuanced perspective on these conclusions. The general trend that can be extracted from them is that BA 45 is specifically involved in processing contexts that are representationally complex but predictable, while BA 44 and LIFS reflect processing effort more generally.

For example, Santi & Grodzinsky (2010) contrast center-embeddings and long-distance dependencies by repeating sentences of the same kind, with the intention of finding out which regions deactivate during repetitions. This phenomenon, known as repetition suppression, forms part of the tenets of predictive coding: the amount of prediction error is reduced with repetition, as this generates stronger priors (higher certainty expectations) that explain away prediction error before it takes place. Santi & Grodzinsky (2010) find that BA 45 adapted to long-distance dependencies but not to center-embeddings, that no region adapted to center-embeddings but not long-distance dependencies, and that BA 44 adapted to both long-distance dependencies and center-embeddings.

Likewise, Santi & Grodzinsky (2012) test reflexive constructions like “The waitress pinched herself”, which are also considered long-distance dependencies. They do not observe the same association between BA 45 and distance, since the appearance of a reflexive pronoun would not be predictable on the basis of the subject (cf. also Santi & Grodzinsky, 2012).

While these results seem to favor a working memory account in which BA 45 fulfills maintenance functions provided that a top-down prediction needs to be maintained, other results seem to cast doubt on this interpretation.

In particular, Wells et al. (2009) show that the processing of object-first sentences can be trained and consolidated, which speeds up processing times on par with subject-first sentences. Provided that the activation increases associated with object-first sentences were eliminated as a result, this would be problematic for accounts associating working memory with specific regions: the need to maintain predictions in working memory across intervening elements never disappears, but the regions that supposedly handle these processes are no longer sensitive. The study of Santi & Grodzinsky (2010) on repetition suppression can also be interpreted similarly: increasing the certainty of the cognitive system about what type of construction to use eliminates processing costs, increases performance, and mysteriously deactivates regions supposedly associated with movement (Santi & Grodzinsky, 2010) even though movement requirements never cease to exist.

Additionally, BA 45 engagement also seems to be associated with the conceptual identity of the phrases involved in long-distance dependencies. Santi et al. (2015) show that when the intervening phrases in object-first relatives are noun phrases (NP) similar to the dependency head, IFG activity increases, reflecting costs related to the interference between NPs. If the intervening material is a clause boundary, on the other hand, superior frontal gyrus (SFG), another FPN region, engages instead.

While the authors interpret these results in terms of “a grammatically instantiated type-identity (i.e., NP)”, *conceptual identity* (associative semantics, consolidation) may be a better description of what is affecting BA 45 engagement, in light of other studies.

For example, the degree of lexical knowledge in children positively correlates with comprehension of object-first relatives, while working memory capacity does not (Rusli & Montgomery, 2017).

Gennari & MacDonald (2008) show that object-first relatives are processed more efficiently when the object NP is an inanimate object, as in "The article that the senator attacked was retracted". Here, evidence related to the semantic feature of *animacy* seems to be having an effect on syntactic processing (cf. also Paczynski & Kuperberg, 2011).

Also, Nicenboim et al. (2016) show that less fluent readers, with less working memory span, process long-distance dependencies faster (and less accurately) than more fluent readers when an embedded subject is *longer* ("Sofía asked who was that *the younger sister of María* had greeted" vs "The younger sister of Sofía asked who was that *María* had greeted"). They attribute this result to the fact that less fluent readers fail more often to retrieve the contents of the embedded NP fully. Importantly, they note a difference between both groups: fluent readers start to show difficulties earlier in time than predicted, while they are processing the NP and have not encountered the verb yet.

These results should also be compared with Traxler et al. (2012), who show that fast readers are more sensitive to semantic content when building semantic structure, and to Conway et al. (2001), which show that low working memory span individuals are more likely to hear their name mentioned in conversations they are not actively listening to (the cocktail party effect).

Throughout this dissertation we will develop a framework inspired by predictive coding to account for results like these. In what follows, an overview of the model will be introduced, and we will finish this section with a reinterpretation of the results presented so far.

The main underlying assumptions are that factors like content, consolidation and individual domain-general capacity differences always play a role when determining which patterns of activity will emerge when processing linguistic structures, and that such patterns of activity reflect the efficiency of a single processing system in integrating parallel sources of information from different cognitive domains (cf. the *global workspace,* Baars, 1997; Dehaene et al., 1998), rather than different functions like working memory and syntax.

Thus, highly efficient integrations, whether these are coming from possessing a higher cognitive bandwidth or stronger priors (e.g., a richer preceding context, priming effects, the processing of well consolidated elements), should be associated with the capacity to progressively anticipate (pre-activate, simulate) with more specificity the bottom-up input before it arrives. Any source of evidence should be valid to achieve this goal, with the cognitive system having the ability to flexibly focus on them and assign them different degrees of reliability in order to arrive at a coherent interpretation (Kuperberg, 2016).

In neural terms, the proposal is that domain-general cognition is instantiated by a *rostrocaudal hierarchy* for the online representation of increasingly more complex representations, starting in the occipital and parietal lobes and ending in the dorsal bank of BA 45 (Nee & D’Esposito, 2016; cf. also Koechlin & Summerfield, 2007; Badre & D’Esposito, 2009).

In turn, every dorsal level of this hierarchy should possess a higher-level ventral counterpart, representing a *predictive coding hierarchy*, and going from medial temporal lobe (MTL) to orbitofrontal cortex (OFC) across limbic cortical regions (cf. Barbas & Rempel-Clower, 1997; Olson et al., 2007; Barrett & Simmons, 2015; Chanes & Barrett, 2016).

These two hierarchies should not be confused: the rostrocaudal hierarchy represents progressively wider spatiotemporal windows for the integration of widespread, parallel brain activity, whereas the predictive coding hierarchy represents the control of this process via feedback projections. Thus, in this model dorsal activity relays to more ventral regions the difference between prediction and prediction error, acting as the evidence necessary to adjust ventral predictions. Dorsal-ventral integrations would take place constantly, producing a series of *predictive cycles*. Accordingly, any cognitive function should always require an interaction between dorsal and ventral activity.

Crucially, the predictive coding hierarchy can override (inhibit) dorsal activations during conditions in which the cognitive system is highly certain about how to integrate upcoming bottom-up input. In these cases, activity would displace towards ventral regions and reaction times would be faster. On the other hand, if the cognitive system was in a less certain state, the ventral stream would be acting more inefficiently when trying to curb prediction error, so dorsal activity would start spreading laterally and upwards across the rostrocaudal hierarchy. These two states of certainty are known here as a more *proactive* vs *reactive* ventral stream.

Since the present model attributes all dorsal increases of activity to inefficiencies in ventral feedback, the association of serial order with the dorsal stream (Bornkessel-

Schlesewsky & Schlesewsky, 2013) is understood as deriving from the increased integration demands that requires the processing of a sequence in which the order of the elements matters, instead of as a one-to-one correspondence between stream and function. This view is reinforced by the association between the ventral stream and the hippocampal complex via medial temporal lobe, as there is evidence that the hippocampus is able to track serial order (Albouy et al., 2013).

Finally, the cognitive system is able to make efficient use of its metabolic resources, estimating the potential rewards of engaging with the input. This is captured in predictive coding terms by the notion of *precision*, which represents post-synaptic gain of specific populations of error units. When attributed a higher precision, these populations would boost their influence on the higher levels of the hierarchy, driving learning and compensatory action, while suppressing competing populations (Feldman & Friston, 2010; Pezzulo et al., 2015; Auksztulewicz & Friston, 2016).

In terms of this model, this means that there is a flexible focus of top-down attention that can be autonomously displaced towards high-level ventral regions or lower-level dorsal regions, capturing the different activation patterns that can emerge for the processing of the same linguistic structure. It also means that the mobilization of metabolic resources when engaging with a task can be different across individuals and contexts (Ferreira & Patson, 2007; Traxler, 2014).

We can now return to the experiments discussed above and the functional relationship that exists between the FPN and FTN. The main hypothesis is that there are three main patterns of activity applying to all of these cases, representing a continuum of progressive increases in prediction error: domain-specific FTN engagement with uncorrelated FPN activity (discussed in Chapter 2), increased left-hemispheric FTN-FPN synchronization spreading laterally to LIFS, and increased bilateral FTN-FPN synchronization possibly spreading to SFG. In order to determine which general pattern of activity will emerge, it is necessary to know whether the ventral stream is in a more proactive or reactive state, i.e., how much parallel dorsal feedforward activity remains to be integrated after ventral feedback.

The association of BA 45 with movement or the processing of long-distance dependencies (Makuuchi et al., 2009; Santi & Grodzinsky, 2010) follows from the cognitive system being engaged with the second pattern of activity. This should result from a more proactive ventral stream, in which an estimation has been made that the bottom-up input has a high level of representational complexity but can be reliably anticipated with some level of abstraction. This translates into more constrained (left-hemispheric) and anterior (BA 45) dorsal activity, representing via dynamic coding mechanisms (Stokes et al., 2013) a highly multimodal and integrated signal that can simultaneously and indivisibly encompass neural populations belonging to different cognitive and linguistic domains (e.g., number, vision, phonology, syntax, semantics).

This state should only emerge when the cognitive system is making rapid use of most of the available evidence to guide itself towards a reliable interpretation, including statistical or associative content, lexical access, contextual information, long-term memory, etc. This seems to be the reason why BA 45 is specifically associated with the processing of syntactic structures on the basis of meaning (e.g., Santi et al., 2015).

BA 45 engagement should nonetheless represent a moderate degree of uncertainty, as language is too flexible to be fully anticipated, and reliable predictions can normally be made only at a deeper level of abstraction than the word level. Such uncertainty would be represented by prediction error spreading laterally towards more domain-general regions within BA 44, BA 45, and ultimately LIFS.

Thus, factors that decrease this uncertainty, like repetition (Santi & Grodzinsky, 2010) or consolidation (Wells et al., 2009; Rusli & Montgomery, 2017), can explain away prediction error specifically in the highest level of the rostrocaudal hierarchy, the dorsal bank of BA 45. This would entail both a reduction of overall activity and a displacement of the higher-level functions of BA 45 to even higher-level regions of the ventral stream.

For example, if the processing of an object-first relative would require the inhibition of the more consolidated subject-first relative, short-term plasticity could increase the weight of the object-first interpretation, eliminating dorsal competition earlier in time. Then, the nature of the prediction error that would remain would be lower-level, and therefore associated with posterior regions like BA 44. This prediction error would represent the part of the bottom-up input that does not have any consequences at all for the higher-level prediction of whether to instantiate a subject-first or object-first sentence. A specific example could be the actual words appearing in the bottom-up input, if these could not be reliably pre-activated.

Importantly, all of this is not language, but a brain state that allows efficient linguistic processing under conditions of high representational complexity. The states in which a more reactive ventral stream is being deployed should be progressively associated with more uncertainty (prediction error), and as a consequence more overall activation that is more distributed across the brain, more metabolic expenditure, slower reaction times and more effort.

Participants with lower domain-general capacities, less able to coordinate integrations under a proactive high-level plan, would either dedicate less metabolic resources to the task, achieving less representational complexity and accuracy of comprehension (e.g., Nicenboim et al., 2016), or adopt this more reactive, expensive and inefficient strategy, leading to their skipping some sources of information entirely or to integrate them across more serial steps or predictive cycles (Duncan, 2012).

However, this brain state also seems to be necessarily associated with the most efficient way of succeeding in some tasks, including the processing of center-embeddings (Fiebach et al., 2004). Thus, it is likely that the processing of center-embeddings is not associated with the suppression of BA 45 through repetition (Santi & Grodzinsky, 2010) because a threshold of prediction error is consistently reached when processing these structures, rather than because BA 45 is not involved in them. This would produce a partial desynchronization of the dorsal and ventral streams (e.g., Fornito et al., 2012), making the contributions of structures like BA 45 resemble those of BA 44 more than during higher-level proactive conditions.

A hypothesis for future work is that the superior frontal gyrus engagement found during intervening clauses as opposed to intervening NPs (Santi et al., 2015) may be closer to this pattern of activity, representing a larger surge of prediction error that would promote a momentarily less proactive ventral stream. In support of this, lateral SFG has been specifically associated with the highest level of executive processing (Boisgueheneuc et al., 2006), while medial SFG is associated with task switching when a strategy becomes unreliable (Nagahama et al., 1998; Cutini et al., 2008).

In sum, the solitary engagement of the FTN vs the collaboration of the FPN and FTN during different conditions of language processing does not entail that these networks must have fundamentally different functions like working memory and syntactic processing. An alternative account based on the spreading of prediction error towards domain-general regions when constantly required higher-level ventral/limbic predictions fail to account for prediction error can also capture these results. Moreover, such an account is more in consonance with the current consensus in cognitive neuroscience that all processing regions are mnemonic, instead of relying on dedicated storage regions (D'Esposito & Postle, 2015; Eriksson et al., 2015). In the following sections, the mechanisms by which these states are instantiated will be explored more in-depth.

1.4. Event-related potentials as a single processing system

This section examines the different patterns of FPN and FTN activity through the lens of event-related potentials (ERPs), and suggests that both follow a linear function characterized by a process of evidence accumulation in which prediction and uncertainty play a necessary role. This proposal offers a single-stream interpretation of predictive processing based on the combination of feedforward dorsal activity and feedback ventral activity. This section is organized as follows: first, there is a general discussion of the two main evoked responses in language (N400 and P600) while paying attention to how they inform our conceptions of parsing (1.4.1 & 1.4.2); second, the specific cases of a P600 pattern turning into a biphasic pattern when subjects are given more time to process (1.4.3), and of the anterior positivity are explored in order to support the continuity-based perspective offered here; the proposal is then justified on the basis of oscillatory and neurochemical evidence, it is linked with the FPN-FTN relationship, and some experiments are proposed (1.4.4). Finally, there is a summary of the model (1.4.5).

1.4.1. Event-related potentials and prediction in language processing

The two ERPs that are considered here are the N400 and P600 (Kutas & Hillyard, 1980; Van Petten & Luka, 2006; Kutas & Federmeier, 2011; Swaab et al., 2012; Kuperberg, 2013). In language, the N400 occurs after every new word that appears in the input, but its amplitude is larger the more implausible or unexpected the word is according to the ongoing event representation. This effect becomes more significant whenever the cognitive system is able to predict with higher certainty the specific word that is going to appear next (i.e., there is a pre-activation). In those cases, the N400 response is very much attenuated if the pre-activation is confirmed, and there is a significantly more pronounced amplitude increase if it is disconfirmed (Kutas et al., 2014)[4]. Some researchers relate the N400 to the ease of lexical-semantic access (Lau et al., 2008; Brouwer et al., 2012) or to compositional integration costs (Brown & Hagoort, 1993) produced by the current word, which can be influenced by the previous associative context of the sentence. Other researchers put a greater emphasis on the sensitivity of this signal to anticipatory

---

[4] An important reminder on the terminology used here: the words “prediction” and “feedback” refer to top-down input acting on bottom-up input according to the tenets of predictive coding, regardless of whether such input is able to anticipate incoming stimuli or just react to them. The word “pre-activation” refers to a type of prediction in which future bottom-up input is anticipated (pre-activated) and explained away when it arrives. How much the cognitive system can anticipate future bottom-up input (i.e., whether we are talking about a pre-activation or not) is assumed to depend on how constrained ventral predictions are, with the more constrained cases reflecting a probability distribution in which one candidate (e.g., a specific word) has a very high probability of appearing relative to other candidates. Therefore, the distinction between “prediction” and “pre-activation” is gradual.

processes in language, which can be based on multiple scales (orthographic, lexico-semantic, sentential, and discursive) and occur with differing degrees of certainty (Wlotko & Federmeier, 2013; Kuperberg, 2013; Kutas et al., 2014). According to Kutas & Federmeier (2011), the N400 defies the pre- and post-lexical distinction, and refers instead to a time window in which multimodal long-term memory access takes place. This is supported by its localization in different regions of temporal cortex (and occipital and parietal to a lesser extent) (Dien et al., 2010; Khateb et al., 2010; Brouwer & Hoeks, 2013; Gallagher et al., 2014) with parallel contributions from both hemispheres (Wlotko & Federmeier, 2013).

The P600 was initially related to syntax given its emergence for cases of repair and re-analysis of syntactic anomalies like agreement errors and garden path sentences (Friederici et al., 1993; Hagoort et al., 1993; Kaan et al., 2000). It is also present, with a more anterior distribution, right after the processing of well-formed long-distance dependencies (Gouvea et al., 2010). Nonetheless, more recent studies also talk about a *semantic* P600, which indicates that the phenomenon cuts across traditional linguistic boundaries like these. For example, a P600 effect not preceded by an N400 effect is elicited when processing reference attribution (Burkhardt, 2006), more demanding contextual inferences (Burkhardt, 2007), or irony interpretation (Regel et al., 2011). The most famous semantic P600 effect arguably occurs when processing role-reversals, such as “Every morning at breakfast the eggs would eat…” (Kuperberg et al., 2003; Kim & Osterhout, 2005). The source of the P600 is often localized in either BA 44, 45 or 47 (Brouwer & Hoeks, 2013), or anterior cingulate cortex and right anterior prefrontal cortex (Shen et al., 2016). An initial generalization is that the P600 is related to the conditions of increased cognitive control that suppose the engagement of the FPN, described in Section 1.1.

As we can see, the association of the N400 with semantics and the P600 with syntax seems to be an oversimplification. Nowadays, these signals are considered in domain-general terms as indexing “long-term memory activation” (N400) and “decision certainty in cognitive control” (P600) (Dröge et al., 2016). This change is motivated by the N400 being observed for the processing of “drawings, photos, and videos of faces, objects, and actions; sounds; and mathematical symbols” (Kutas & Federmeier, 2011), which has led to its interpretation as reflecting the context-dependent access of semantic features, or, in predictive coding terms, as a signal that marks the online construction of the causes of the ongoing event (Kuperberg, 2016). The N400 also appears for some syntactic conflicts, especially in languages in which these constructions are less order-dependent (Bornkessel-Schlesewsky et al., 2011). Likewise, the P600 has been shown to be sensitive to violations of non-linguistic sequential order (e.g. Lelekov-Boissard & Dominey, 2002; Núñez-Peña & Honrubia-Serrano, 2004), and is argued to be a variant of the P3, an earlier ERP that is sensitive to factors of probability, salience, and task relevance beyond the language domain (Coulson, 1998; Bornkessel-Schlesewsky et al., 2011; Dröge et al., 2016).

The domain-generality of the ERPs and their association with predictive processes has consequences for our conceptions of parsing, as they enable the proposal of an alternative to two-stream hypotheses that are based on syntax and semantics operating with different time courses. An example of a two-stream hypothesis is found in Kim & Osterhout (2005) to explain the P600 that occurs for role-reversals. The main idea is that sentences like “Every morning at breakfast the eggs would eat…” elicit a P600 because there is a conflict between a faster stream that processes associative semantics and a slower stream that pays attention to the superficial syntactic structure. The *semantic illusion* that “eggs are eaten”, based on the subject’s world knowledge, would be attained first, and the slower syntactic parser would then trigger a P600 at a later stage, signaling that the syntax of the sentence does not match the intended meaning. These conclusions bode well with analysis-by-synthesis theories of language comprehension, which divide processing into two stages: first, a surface-to-semantic representation “that is (at least close to) correct” is created. Then, “a derivation that is correctly directed to generate the input surface form” is enacted (Bever & Poeppel, 2010).[5]

Offering an alternative, one-stream view, Chow & Phillips (2013, cf. also Stroud & Phillips, 2012) defend that the semantic analysis proceeds simultaneously with the superficial syntactic structure of sentences during language comprehension. They note that sentences with no possible role reversal and that are syntactically well-formed (e.g., “The student hung the math problem”) still elicit a P600 effect (although less positive and preceded by an N400 effect), showing that the introduction of two streams with different

---

[5] The approach may be additionally used to hypothesize on the neural reality of right-branching derivational orders, which are closer to formal linguistic theory (e.g., Chomsky, 1995/2014) (Phillips & Lewis, 2013). The main adjustment needed is a substitution of the “surface-to-semantic representation” for this formal machinery, which would be responsible for building the initial phrase structure on which the second stage, re-baptized as “linearization”, is based.

time courses is justified by their ability to account for a subset of the data only, and that boundaries like these need to be based on (or be compatible with) neural distinctions.[6]

The main neural distinction that still seems to justify two-stream accounts is that of the ventral and dorsal streams, which seem to be the seat, respectively, of the N400 and P600 responses given their proposed sources. This association is reinforced by experiments like Lau & Nguyen (2015), where the researchers altered the timing of the screen appearance of a predicted (pre-activated) lexical item, so that time-sensitive [dorsal] predictions were more or less effective across conditions. They found that the effectiveness of temporal predictions did not help the effectiveness of semantic predictions about the word that was due to appear, and that the N400 amplitude did not change across conditions, which suggests that the lexical pre-activations were on a time-insensitive [ventral] format.[7]

However, experiments like these do not offer any conclusive evidence for each stream being able to operate entirely in parallel, just showing that it may be possible to distinguish processes that require a more extensive or pronounced ventral contribution from others that do not. In other words, language processing may exploit both streams

---

[6] Regarding the absence of an N400 during the semantic P600 in role-reversals, Chow & Phillips (2013) introduce the hypothesis that it is due to properties related to the surface syntactic structure, specifically to the target verbs being lexically accessible, due to a combination of "strong lexical associations", and "weak contextual constraints", as opposed to the cognitive system being sensitive to the presence of a plausible non-surface structure. The P600 response is not functionally interpreted in these studies, but cf. Brouwer et al. (2012) for a compatible account based on integrative processing, or "difficulties in establishing a coherent utterance representation" (Brouwer et al., 2017).

[7] Lau & Nguyen (2015) also claim: "lexical-semantic predictions may be maintained over a broad time-window, which could allow predictive facilitation to survive the presence of optional modifiers in natural language settings". For example, a pre-activation like "it was sunny outside so the boy went outside to fly a… kite" would not be invalidated by the appearance of "a *new* kite". This agrees with the account of non-adjacency of Section 1.3., where it was claimed that there is a necessary role for the ventral stream to establish these relationships despite their association with increased dorsal activity.

rapidly and interactively, with more complex operations (like the online generation of a pre-activation that is maintained) requiring increased dorsal activity before the ventral stream can solve the input and proceed to the next operation. This unequal relationship, in which the dorsal stream would provide feedforward signals (prediction error) and the ventral stream feedback signals (predictions that explain away dorsal stream activity), allows for a unification of two- and one-stream accounts, provided that the identification of each stream with syntax and semantics is abandoned.[8] Therefore, the fact that there are two physical streams associated with language does not make one-stream accounts of parsing necessarily indefensible, but can add information to their functioning.

In short, the premises adopted by the present account are the following: 1) that past processing (context, world knowledge, consolidation) influences current processing, meaning that purely bottom-up processing does not exist, that top-down influences can be generated at multiple scales and with differing degrees of specificity and reliability, and that the timescale of processing is not completely fixed but can vary according to the context, individual capacities, and task demands; 2) that the dorsal and ventral streams do not map on a one-to-one basis with syntax and semantics nor act independently, but work interactively and cyclically on the linguistic input for a variably prolonged duration; and 3) that the ERP components and the different patterns of FPN-FTN interactivity do not constitute fundamentally different functions if examined with sufficient granularity, but

---

[8] One-stream views are also compatible with analysis-by-synthesis accounts, in the sense that there should be a gradual accumulation of evidence happening over the received input and associated with increased effort (progressively building a more elaborate representation), but with the additional possibility of interrupting this process earlier if the ventral stream can solve the input by appealing to long-term memory representations. This would reduce representational complexity in favor of speed.

arise from a single mechanism performing with differing degrees of efficiency and anticipatory capacities.

Given that, the main obstacle of one-stream accounts like this one regarding the ERP literature seems to be in explaining why we find patterns of evoked activity that are so suggestive of different processing mechanisms activating under different conditions (i.e., the semantic P600 with no N400, the N400, and biphasic patterns). The answer provided here is that there are only two related factors involved in all of these different patterns of activity, which are how certain the cognitive system is about upcoming input at any particular moment (in other words, how constrained ventral predictions are), and how well those predictions fare when the input appears.

Evidence for the continuity of the ERP signals (vs each signal being the product of one stream operating independently) can be found in multiple sources, reviewed in what follows. After this brief review, the case of the semantic P600 will be explored in more depth and rephrased to account for the present proposal.

1.4.2. Evidence for contextual variability in ERP patterns

There are multiple sources of evidence in which the ERP change for the same task depending on the processing context. This has forced the conclusion that the brain is capable of using different processing strategies, but the neural basis for these is often localizationist, with specific regions or streams being singlehandedly responsible for their deployment. An alternative way of looking at the issue is considering that quantitative measures like efficiency, certainty, degree of consolidation, individual capacities, etc.

may change across subjects and trials, changing in turn the extent and locations in which prediction error accumulates, and the way in which the ventral and dorsal streams engage. The sources of evidence for this one-stream view can be categorized in at least four groups:

1) *Individual differences*. For example, domain-general cognitive control abilities measured by the Stroop task are associated with different response profiles. The processing of implausible active sentences elicits a negativity in the best performing group, but a sustained positivity (350-750 ms) in the worst performing group (in non-canonical, passive sentences a sustained positivity is observed in both groups) (Ye & Zhou, 2008, 2009). Nakano et al. (2010) show what seems like the opposite pattern, with the sentence “The box is biting the mailman” eliciting an N400 at the verb for low working memory span subjects, but a P600 when working memory span was higher. Also, second-language learners initially show an N400 effect for subject-verb agreement violations, which get replaced by a P600 response as proficiency increases and item-based representations shift towards more generalized, rule-based processes (Osterhout et al., 2006; McLaughlin et al., 2010; Tanner et al., 2013; Tanner, 2013).

Two-stream hypotheses focus on the idea that each stream is responsible for a different strategy, one based on combinatorial aspects and the other based on long-term memory representations (Kuperberg, 2007), usually with no stream showing any primacy over the other in general, even if some constructions can be processed more efficiently by one of them, or higher domain-general capacities may prompt an individual to use one over the other in some contexts (Tanner, 2013).

An alternative, one-stream view is that each strategy simply reflects precision shifting (Feldman & Friston, 2010; Pezzulo et al., 2015; Auksztulewicz & Friston, 2016) occurring from encountering and focusing flexibly on diverse cues or sources of evidence (word order, animacy, thematic relations, event representations…) (Kuperberg, 2016). These cues vary in how order-sensitive, easily processed, consolidated or capable of anticipating incoming input at a particular context they are, leading to a ventral stream that is more or less helpful at reducing prediction error coming from the dorsal stream. While both accounts seem to be equivalent for the most part, they make different predictions regarding the independence of the dorsal stream, with the present account suggesting that both the beginning and end of a period of increased dorsal activity is marked by a transitory increase of ventral activity, signaling that the dorsal stream may be a slave mechanism instead of an independent processing route. A similar proposal to this is that "language processing proceeds in N400-P600 cycles" (Brouwer et al., 2012).

2) *Cross-linguistic differences*. For example, role-reversals have been shown to cause a semantic P600 in English and Dutch, a biphasic N400-P600 pattern in German, and an N400 in Turkish and Chinese (Schlesewsky & Bornkessel-Schlesewsky, 2009; Bornkessel-Schlesewsky et al., 2011). Two-stream approaches explain this discrepancy by suggesting that some languages are generally more order-sensitive than others. The lack of an N400 effect seems to occur in languages with more rigid word orders, in which thematic relations are specified on the basis of ordered information alone. As word order is the only source of evidence taken into account by English speakers, there is no possible conflict with other sources (e.g., case, animacy) at the time window of the N400, in which thematic relations are supposed to be processed. The authors suggest that this entails the

use of different processing strategies “which may result in cognitively and neurobiologically distinct processing mechanisms” (Bornkessel-Schlesewsky et al., 2011), presumably the dorsal and ventral streams (cf. Bornkessel-Schlesewsky & Schlesewsky, 2013).

This two-stream approach (the Extended Argument Dependency model, cf. Bornkessel & Schlesewsky, 2006), in which the semantics and syntax distinction is abandoned in favor of operations that are more dorsal or ventral in nature, is in very minor conflict with a one-stream account like the present one, which assumes that the different processing strategies are different ways in which dorsal-ventral interactions take place (i.e., precision shifting). The assumption that different languages may prime the cognitive system differently when facing the choice of where to spend more processing resources can be accommodated as well by the notion of precision.

However, the main point of contention has to do with the processing of thematic roles, which is assumed to always occur at a second step of processing in which relations between the verb and its arguments are computed, having consequences for the N400 (Bornkessel & Schlesewsky, 2006). The alternative proposal discussed in the next section is that the lack of an N400 effect for role reversals when word order is the only source of evidence may hint at the possibility that thematic relations can have a delayed impact on ventral predictions (and therefore no N400 effect). This follows from a one-stream conception of the parser, because it would be a case in which a (relatively slower and more effortful) increase in dorsal activity is a requirement for the ventral-related N400 to be affected. It also assumes that the order of operations is not necessarily fixed

(Kuperberg, 2016), but that, based on domain-general principles, some operations can be faster or easier in the absence of any strong priors (pre-activations).

3) *Token-related differences*. There are multiple cases in which different tokens of the same type tend to evoke different responses, affecting the observed components across the board. For example, Kim & Osterhout (2005) detect a monophasic P600 for “The hearty meal was devouring”, but Kim & Sikos (2011), while replicating the results for the same sentence, find that the different token “The hearty meal would devour…” elicits instead a left anterior negativity at 300-600 ms. Similarly, Kim & Osterhout (2005) report a biphasic N400+P600 for the sentence “The dusty tabletops were devouring…”, but Kuperberg (2008) observes a monophasic P600 for “Every morning at breakfast the eggs would plant…”. Kuperberg (2007) attributes these differences either to component overlap, or to the mechanism behind the P600 being able to “switch off” the mechanism behind the N400 in some cases. While the latter explanation is thought out in terms of a two-stream hypothesis, it is in consonance with a one-stream hypothesis provided that *switching off* means that sometimes the higher levels provide a very coarse prediction of the activity in the lower levels, and so they are not affected by small differences in the lower levels.

This explanation also seems to be what is going on when contrasting a sentence like “The student restrained the math problem” with the very similar “The student hung the math problem”. While the former elicits a moderate N400 effect followed by a moderate P600 effect, the latter elicits a comparatively stronger N400 effect followed by a much smaller P600 (Chow & Phillips, 2013). The authors interpret this discrepancy based on the fact that the first sentence has a verb that is animacy-incongruous, increasing the P600 signal. Since they interpret the N400 as a marker of ease of lexical access in which animacy should not have any impact, they attribute the N400 effect to either component overlap or some unaccounted difference in the ease of lexical access of the verbs “restrained” and “hung” that would (not intuitively) favor “restrained”.

The alternative view, following our reinterpretation of the “switching off” idea, is that animacy-congruous verbs (“hung”) tend to favor a higher precision on the ventral stream, reducing P600 effects while augmenting N400 effects. Thus, the relevant point of comparison between both sentences is not how much the verb is facilitated (as indexed by a smaller N400), but to what extent it can have a significant impact on the ventral stream. In other words, in conditions that are unpredictable (no strong priors), the P600 may be acting as a barrier that reduces the effect a word can have on the N400, regardless of how accessible it is. In this case, the reason would be that the animacy-incongruity of the verb “restrained” forces the system to switch precision towards the dorsal stream. This hypothesis is consistent with a one-stream account in which the ERPs are interrelated.

4) *Latency-related differences*. Chow (2013, Chow & Phillips, 2013; Chow et al., 2015, 2016) looked at the N400 amplitude to distinguish whether a target verb was facilitated on the basis of event representations gathered from the thematic roles fulfilled by its arguments. For example, in the sentence “The cop arrested the thief” (following a SOV order in Mandarin), the verb “arrest” should be facilitated by “cop” being the agent and “thief” being the patient. They found that this is not the case, because when contrasting this sentence with the role-reversed “The thief arrested the cop” a similar N400 amplitude is present (with the latter engendering the well-known semantic P600).

However, when a short time buffer or intervening words are introduced between the arguments and the verb, the role-reversal engendered a biphasic N400+P600 pattern.

This suggests, according to the authors, that event representations take more time to be constructed than facilitatory effects due to semantic associations, which seem to be present from the start. Importantly, this latency difference cannot be attributed to different streams according to the perspectives studied before: the researchers show that the arguments of the verb, thought to be syntactic, are processed from the very beginning, as the associative facilitatory effects seem to be dependent on their identity as arguments of the verb. This hypothesis is known as the “bag-of-arguments” hypothesis, in contrast with the “bag-of-words”, in which syntactic information (hierarchy) does not bear on processing until a slower syntactic stream engages later.

In sum, all the cases studied have in common a variability of responses that is not strong enough to prevent making some generalizations about the parser, the main one being the existence of two strategies for dealing with incoming input that are flexibly established depending on multiple contextual factors. It can be argued that these strategies are not uniquely associated with each ERP or stream, but are distinguished by how constrained ventral predictions are at any particular time, or how precision can shift towards the lower levels of the hierarchy when difficulties arise in predicting the input. Thus, significant N400 responses become more likely the more constrained ventral predictions are or the more precision is placed on the higher levels, and P600 responses become more likely when more precision is placed on the lower levels, which can be due either to mispredictions (pre-activations that are disconfirmed), or to unpredicted input.

The next section considers the phenomenon of unpredicted input from the point of view of the latency differences observed for role-reversals, and the following one contrasts it with disconfirmations as indexed by anterior positivities. In doing so, it defends that the internal structure of prediction is fixed according to neural parameters, but not the time course of events, which can change according to which strategy is chosen and how efficient it is for dealing with the specific input.

1.4.3. Events as dorsally constrained ventral predictions

The interpretation favored by Chow et al. (2016) for the latency differences observed during role-reversals is that syntax and semantics are computed simultaneously (one-stream hypothesis), but event representations (i.e., the N400 becoming sensitive to compositional semantics or thematic roles, and not just semantic associations) take longer to be generated. More specifically, the proposed order of operations is the following: first, there is a "bag-of-arguments" mechanism that identifies the arguments belonging to a verb and derives semantic associations based on them, which are used to generate multiple candidate events in long-term memory. If the subject is given more time to process, the second mechanism engages, in which these candidate events are serially contrasted with the linguistic cues in order to find the right match, a process that takes relatively more time and effort. If processing is interrupted before the second step takes place, a monophasic P600 emerges for "The thief arrested"; if it is interrupted after it, a biphasic N400+P600 pattern emerges, perhaps indicating that accessing "arrested" is costlier.

A possible reason for the existence of this delay or second step, they suggest, is that there is a mismatch between the format of linguistic cues and long-term memory. To illustrate this difference, they propose the contrast of long-term memory representing events in the more specific “cops-as-arresters”, while linguistic cues would lead to the more abstract “cops-as-subjects” or “cops-as-agents”. This interpretation is different from the ones discussed in the previous section because it advocates for the simultaneous processing of syntax and semantics but at the same time acknowledges intrinsic time differences in the operations that the parser performs.

In her commentary of these papers, Kuperberg (2016) shows how predictive coding and one-stream accounts can try to accommodate these findings. This author suggests that the idea of a semantic illusion can be useful to explain them, but not because of the existence of a faster non-surface semantic stream. Instead, the argument is that a cognitive system derived from an actively generative predictive coding mechanism does not always have to focus on all the available information in order to make pre-activations about the upcoming input. Focusing on a subset of the evidence becomes possible when this subset is judged to be more reliable (is weighted more) than the full set (the full interpretation of the linguistic evidence). For example, in the role-reversed “The restaurant owner forgot which waitress the customer had served” (Chow et al., 2015), Kuperberg (2016) suggests that the linear order of “waitress” followed by “customer” is used to pre-activate a familiar restaurant schema that promotes the event <waitress served customer>. Then, a smaller N400 emerges not because there is no conflict with long-term memory, but because the pre-activated semantic features deriving from that event explain away the prediction error of incoming words that match those pre-activations, reducing the overall response. Then, as more time passes and the syntactic-thematic properties of “served” become available via bottom-up input, the initial interpretation turns around. In other words, semantic illusions emerge during the processing of cues that are so reliable that they manage to temporarily shift attention or precision towards the events they suggest, as opposed to the full interpretation of the actual sentence.

This phenomenon of precision shifting is different from the one proposed previously to explain the different processing strategies, because it does not move between high and low levels of the hierarchy, but between potential cues that are promoted over others for further interpretation, which could be encoded in parallel outside of the focus of attention. Nonetheless, both processes are likely to form part of the conception of parsing that is being defended here, as they have in common that processing does not need to proceed in exactly the same manner for all contexts and subjects, as this would be inefficient given the possibility of using long-term-memory-based predictions (consolidation) to immediately and effortlessly account for the incoming input.

However, this account still leaves important questions unanswered. In their response to Kuperberg (2016), Chow et al. (2016) state that an explanation is still needed for why the cognitive system should shift consistently and across different sentences from relying first on bag-of-arguments (linguistic) cues to relying on argument + role (thematic) cues after more time passes, unless their original assumption that argument + role cues are in a different format from long-term memory holds. This effect is also shown cross-linguistically, even in languages in which argument + role cues are not encoded by word order alone (for example, by case markers in Chinese).

The argument that will be adopted here is that, in the absence of strong priors, there is an intrinsic (cross-linguistic, universal) difficulty in generating pre-activations based on relational information, which explains the observed latency differences and ERP patterns. This difficulty does not have anything to do with the accessibility of argument + role cues (i.e., these cues can be present and taken into account by the parser from the start), but with their use for generating pre-activations: the cognitive system has a natural tendency to predictively use semantic associations first, because making structural pre-activations (or pre-activating a verb based on them) takes more effort.

The reason why it takes more effort is that structural pre-activations are both very specific and informationally dense: they represent the construction of a single model of future input based on multiple participants holding a specific relationship between them (i.e., not all automatic semantic associations are valid), which is held with great confidence even if there is an absence of strong bottom-up evidence in its favor. This can also be interpreted as the simulation of events requiring a higher spatiotemporal representational resolution, and therefore the engagement of higher levels of the cognitive control hierarchy at the expense of the processing speed conferred by a proactive ventral stream.

The difference of this account with Chow and colleagues' is that these researchers suggest an additional operation in which argument + role cues (e.g., thief-as-agent) are serially contrasted with a set of verbs that were pre-activated in parallel based on bag-of-arguments cues alone, a process that takes time. The result of this additional operation would be the generation of a pre-activation that takes into account the thematic role of the arguments. The existence of this extra step is derived from the idea that event representations in long-term memory, presumably a requisite for these structural pre-activations, are in a non-linguistic format. This entails that linguistic cues and events cannot be directly mapped onto each other, requiring instead a more time-consuming serial search to pick out the candidates that match the linguistic input.

On the other hand, the present account suggests that the online generation of structural pre-activations (also known as the simulation of specific event representations) is an intrinsically costly process in domain-general terms (i.e., regardless of whether the subject is engaging in the processing of linguistic cues or not), so the chosen processing strategy when comprehending is to remain neutral and wait, shifting precision towards the lower levels of the hierarchy. This strategy is optimal even for very predictable events (according to offline measures like cloze probability) because the appearance of a target verb always influences strongly any event simulation, greatly reducing its processing costs by offering reliable evidence that helps discard competing alternatives. Thus, the more time it passes without this strategy yielding a reward, the more likely it will be that the cognitive system will shift towards a more proactive strategy and spend the resources on trying to simulate (pre-activate) an event, explaining the latency effects. From this also follows that the observed effect may not be due to incompatible formats between the linguistic signal and long-term memory, but simply to a continuum in which semantic representations deriving from the linguistic signal, already present from the very start, become progressively more higher-level (chunked, Miller, 1956).

Evidence for this proposal can be found in experiments like Kukona et al. (2011), in which it is shown that pre-activations based on semantic association instead of thematic roles seem to dominate anticipatory looks towards pictures representing upcoming nouns.

In particular, subjects perform nearly at chance when hearing “Toby arrests the…” and having to look at either the picture of a crook or a policeman, even though the agent role is already filled from the onset of the sentence. Nonetheless, structural predictions still have a slight advantage over priming in this study, and the authors claim that their results are compatible with both two-stream (Kuperberg, 2007) and one-stream models (Tabor & Hutchins, 2004; Tabor et al., 2004). Complementing this study, Gambi et al. (2016) show that, at least for well-formed, consolidated active sentences in which associations are non-informative, structural pre-activations do indeed happen, but, crucially, the rate of presentation in this study is slower than the one used by Kukona and colleagues. In sum, it seems that during the first moments of processing, in the absence of strong contextual constraints and within a local context (the bag of arguments), structured event representations in long-term memory either are inactive, or are weighted equally and need to be disambiguated.

From this account a series of predictions about brain activity follow. If the simulation of an event in the absence of strong priors is a naturally costly process, it should entail increased dorsal activity, associated with more effortful conditions. This is what seems to happen during language production, associated with the engagement of the FPN (Fedorenko, 2014), possibly because production does not have the luxury of relying on bottom-up biases (low-level auditory input) that help discard competing outcomes. Thus, the delayed condition in which the N400 effect re-emerges should be associated with more dorsal activity than the condition with no delay. Given that more dorsal stream engagement is associated with serial operations and not just with effort (Bornkessel-

Schlesewsky et al, 2013), this observation would also be compatible with Chow and colleagues' proposal.

Also, if the event was previously simulated and the context still holds (i.e., we already know who did what to whom from a previous sentence), the simulation of the events suggested by the new sentence will be more biased and take less effort, which will be reflected by more chances to find an N400 effect for the non-delayed condition. This may be the time when increasing higher-level precision may be the optimal processing strategy, and may suggest that the higher-level ventral stream is reacting earlier than expected compared with conditions in which no previous context is offered for the processing of the same sentence. A similar paradigm is explored by Federmeier & Kutas (1999; Kutas & Federmeier, 2011), which show that the self-paced reading of the sentence "They wanted to make the hotel look like a tropical resort" followed by the word-by-word presentation of "So, along, the driveway, they planted rows of… [palms/pines/tulips]" changes the N400 response to be uniquely affected by the presence of the preceding sentence, becoming graded from "palms" (the smallest), to "pines", to "tulips" (the largest).

The fact that the N400 changes as a function of the preceding context suggests that the N400+P600 pattern may be recoverable in Chow and colleagues' findings if a similar paradigm is followed in which argument + role cues are bound to a long-term memory representation. For example, in "The thief managed to get a job at the police station" followed by "The thief arrested the cop", the time buffer between the arguments and the verb may not be needed anymore.

Lastly, the P600 that appears for both the delayed and non-delayed conditions is interpreted by Chow and colleagues as an index of semantic plausibility (cf. Brouwer et al., 2012) that is only contextually engaged, but a more complete account follows from the present proposal. The idea is that the P600 should emerge whenever precision is shifted towards more posterior, dorsal regions (lower levels of the rostrocaudal hierarchy). This shift may occur as a result of a pre-activation failing with no obvious alternative to replace it, which would trigger a search for more evidence, but also during those conditions in which the input is unpredicted (pre-activations are not specific enough to account for all the information) and the event needs to be generated from a less biased beginning (in the case of an event representation, from semantic networks that are still very much unconstrained) in order to fulfill the task.

The latter should suppose a monophasic response, whose optional nature (it seems to only be present for one third of all semantic incongruities surveyed by Van Petten & Luka, 2012) derives from the fact that the cognitive system may fail to allocate more precision to the lower levels depending on contextual factors like persisting on higher-level strategies (which may be optimal or suboptimal), committing to an alternative interpretation faster, or even failing to engage attention by switching to an unrelated task.

However, if the cognitive system effectively shifts precision towards the lower levels, the P600 emerges because there is an increase of dorsal activity reflecting the increased need to disambiguate or inhibit competing alternatives represented in parallel, in this case the actual linguistic input versus the non-role-reversed sentences suggested by the subject’s world knowledge. From this, it follows that if a role-reversed sentence was devised so that some subjects did not have the world knowledge necessary to detect the semantic incongruity, they may not show a significant P600 compared to control sentences, but, importantly, they should still show some increased dorsal activity compared with condition that includes a context that primes a specific thematic interpretation. This would be so because, according to the present proposal, the event would still need to be generated with more effort due to the lack of sufficiently specific pre-activations. Since this would be less effortful than adding on top of it a competing alternative that is highly weighted (the non-role-reversed sentence), the total dorsal increase should be less than when world knowledge is present, but still exist.

It should also be noted that the monophasic P600 would not be a dorsal response only: the construction of the role-reversed condition is not independent from the ventral stream but needs to struggle against competing event representations that become instantaneously available as an attempt is made at constraining the associative network in the direction disposed by the bottom-up input.

In sum, the biphasic response may represent a revision of a pre-activation, whereas the monophasic response may represent an unpredicted condition. Crucially, both cases can be understood from the combination of both dorsal and ventral activity, with the only difference being how constrained pre-activations are when the target word appears. This makes the ventral stream feedback-related and necessary in both cases, making the one-stream and two-stream division of parsing theories a simplification: we may be talking about two streams within a single hierarchical cognitive system, with interactions between them being mediated by precision changes.

Crucially, the ventral stream and the N400 can be sensitive to event representations or thematic structure, but in the absence of any strong priors during language comprehension thematic structure may remain unpredicted given the very specific and complex relationships that it entails. This does not mean that the cognitive system is incapable of pre-activating it, but that it tends to make the most efficient use of metabolic resources, as relying on the bottom-up input for its interpretation should produce accurate and fast interpretations at a reduced expenditure compared with simulating an event from higher levels of the hierarchy.

The next section examines a response that has not been discussed so far, known as the anterior positivity, in order to set the basis for linking the ERP literature to the FPN-FTN relationship, and it also offers oscillatory and neurochemical evidence for the proposal.

1.4.4. The anterior positivity; oscillatory and neurochemical insights into the FPN-FTN relationship

The ERPs can inform the FPN-FTN relationship mainly because both are neural realizations of uncertainty occurring at different levels of analysis and studied with different techniques. Nonetheless, the connection is not a direct one: the N400 and P600 have sources that are often beyond the simple division established here, according to which the FPN is bilateral and frontoparietal and the FTN is left-lateralized and frontotemporal. The easiest route of linking the N400 with the ventral stream of the FTN conflicts with the fact that the response has been found across both hemispheres, possibly subserving different functional contributions (Wlotko & Federmeier, 2013). Likewise, positivities are found across cortex, both in regions traditionally attributed to the FTN like Broca's region or posterior temporal lobe, and within the FPN's anterior cingulate cortex.

Nonetheless, a productive way of tackling the problem of linking both neural realities is to explain the fMRI patterns as general long-term consequences of the different ERP patterns. In order to finish this account, we must first explore a positive response that, in contrast with the P600, occurs in anterior regions. This will offer us a chance to explore the FPN-FTN relationship with the support of oscillatory and computational perspectives on uncertainty.

The anterior positivity (AP) seems to emerge for cases of unexpected but plausible scenarios in which upcoming words fit to some degree with the preceding context, without requiring a complete revision of expectations, or in which the bottom-up evidence can be immediately integrated without any effort. For example, Federmeier et al. (2010) prime subjects to expect a type of bird, and match those expectations with the words "robin" (a prototype), the less prototypical "chicken", and the incongruous "potato". Only the second is associated with an AP, while the third elicits a posterior P600. DeLong et al. (2014) observe the AP under the following context: "For the snowman's eyes the kids used two pieces of coal. For his nose, they used a banana", and point that the updates signaled by this ERP have to do with an individual's world knowledge. Wlotko & Federmeier (2012) note that the phenomenon emerges when there are fewer alternative completions than when there are many, and also observe it when presenting an unexpected synonym of a pre-activated word. Kim & Lai (2012) see it with pseudowords that are orthographically similar to preactivations, such as in "She measured the flour so she could bake a *ceke*". Lastly, Gouvea et al. (2009) observe the AP for long-distance dependencies and distinguish it from the posterior P600 related to reanalysis, which supports the idea introduced in Section 1.3 that a ventral prediction plays a necessary role in non-adjacency conditions, with intervening words not being able to disrupt it.

It can be argued that these cases are not qualitatively distinct from those that trigger an N400 + posterior P600 pattern, and that the main difference lies in the amount of cognitive search or disambiguation (i.e., cognitive control) that the revision process entails once a violation has been detected, which usually depends on how appropriate the initial prediction keeps on being after the onset of the target stimulus. If that is the case, it is possible to establish a continuum for all of these responses that can be associated in turn with the different patterns of activity of the FPN and FTN, such that all the observed activity depends on the degree of certainty (how constrained ventral pre-activations are before the input) that the cognitive system has at any particular moment. More specifically, the relationship that is proposed is the following: the more constrained ventral pre-activations are (i.e., the more certain the cognitive system is about upcoming input), the more likely an attenuated N400 response and decreased dorsal activity will be during integration; also, the less adequately ventral pre-activations can account for the input (the more the probability distributions need to change), the more activity will be displaced towards dorsal stream regions.

Oscillatory evidence can further illuminate these views. Piai et al. (2016) compares weakly and strongly constraining sentences in which pre-activations become possible, and notice that the latter are characterized by a relative pre-stimulus increase in hippocampal theta power as measured by implanted electrodes. Rommers et al. (2016), in turn, observe that for strongly constraining sentences there is an alpha power decrease in frontal and central neocortical electrodes, a beta decrease at left posterior and anterior electrodes, and finally an alpha decrease at occipital and central electrodes. These oscillatory changes seem to be the signature of a proactive ventral stream applying its influence over the neocortex prior to stimulus onset.

Also, right after the stimulus appears, strongly constraining contexts are associated with a broadly distributed theta power increase at 300-700ms at posterior, anterior and left medial recording sites, whereas in the same sites a similar alpha and beta power decrease occurs for both strong and weakly constraining contexts at 300-500ms (Rommers et al., 2016). These different oscillatory profiles may mark the distinction between a proactive and reactive ventral stream when it comes to processing, or, in other words, how rapidly the contents of working memory are being updated with feedback from long-term memory. The lack of a theta effect for unconstrained sentences may indeed be indicative of a more local (alpha-based), less integrated, more reactive way of processing, in which integrations take place progressively after the onset of the target word.

The model discussed here can also illuminate part of these results, and makes predictions about the response profiles of contexts that have not been tested so far with these methods. Rommers et al. (2016) note that, for the highly constrained sentences that are disconfirmed by the target word, an N400 effect followed by an AP is elicited. This AP seems to overlap but not completely coincide with the theta power increase at 300-700 ms after the target word’s onset mentioned before. The order of events after word onset is the following: first, a less frontal theta increase precedes a stronger frontal theta increase, which correlates with the AP, both in timing and strength (cf. Sections 2.2.2 and

2.2.4 for the possibility that anterior temporal lobe and ventromedial prefrontal cortex may be behind these results).

Interestingly, the more alpha power decreased before word onset for highly constraining sentences, the weaker the theta increase and late positivity were after the word disconfirmed the pre-activation. This seems contradictory because the larger alpha power decrease could indicate the presence of a stronger or more constrained pre-activation, which would have higher chances to collapse once the bottom-up input appears and does not match it, requiring a more extensive repair process in the form of the AP. However, if the alpha decrease reflects an increase of activity (cf. Lisman & Jensen, 2013) belonging to multiple competing candidates (a graded pre-activation, in which a larger decrease entails a less specified or constrained pre-activation, and therefore more competition), and a long-term memory updating process after word onset works in the longer-distance theta code (Buzsáki & Draguhn, 2004), selecting one of those candidates immediately (hence the AP), these results would make sense. In short, when the pre-stimulus alpha decrease is smaller, it is because the process of coming up with a pre-activation has ellicited fewer competing candidates, or a candidate that is much more highly weighted and specific than others has emerged.

Thus, the expectation would be that in cases in which an N400+Posterior P600 pattern is elicited (not tested by Rommers and colleagues), compared with the N400+Anterior positivity observed here, the pre-stimulus alpha decrease would be comparable or larger. Also, post-stimulus, theta frequencies may be larger, as they would communicate a larger prediction error to the hippocampus and a greater need to exert more cognitive control (Cavanagh et al., 2011; Cavanagh & Frank, 2014). Regarding the

500-700 ms alpha/beta power decrease, reflecting more activity, it could be smaller than in the AP case, signaling the larger inhibitory effort that a P600 would presumably entail.

Regarding the monophasic P600, the differences with Rommers and colleagues' lowly constrained sentences with unexpected outcomes (in which a large N400 effect but no P600 is observed) could be less significant pre-stimulus hippocampal theta and neocortical alpha/beta decreases (signaling a very unconstrained ventral stream), and, after word onset, and a larger alpha/beta power decrease at 400-800 ms (reflecting a larger inhibitory effort).

Lastly, if compared with the N400+Posterior P600 pattern, this last step may be weaker for role-reversed sentences, as the competition should happen between two possible interpretations only. Thus, if the semantic P600 is indeed related to unpredicted as opposed to mispredicted stimuli, it may be possible to show that this alpha/beta power decrease can vary independently from anterior alpha/beta power decreases under different sentences that trigger the semantic P600 (e.g., role-reversals vs non-literal meaning comprehension).

This section finishes by linking the patterns of activity induced by the FPN and FTN with computational models about the neurochemistry of uncertainty.

The anterior and posterior positivities have also been related to the domain-general notion of uncertainty via neurochemistry in computational models, and the same can be done for the FPN and FTN. More specifically, Kutas et al. (2014) relate the anterior positivity with *expected uncertainty*, and the posterior positivity with *unexpected uncertainty*. The idea is that the anterior positivity can be understood as a change in the balance from more top-down towards more bottom-up processing without abandoning the current goal or task set (which explains why we also observe it for pre-activations that are not entirely disconfirmed), whereas the posterior positivity refers the displacement of resources towards the interpretation or the implementation of a new goal or task set, a process that requires both the inhibition of the current set and of competing candidates (hence, enhanced posterior activity and cognitive control).

The notion of expected uncertainty or exploitation refers to conditions in which the cognitive system is unable to predict with accuracy all the contingencies of the expected input, which is a natural consequence of the ever changing environment in which the brain has evolved. During expected uncertainty, stimulus-response-outcome rules from previous tasks can predict future input weakly, but such unreliability is taken into account and remains stable throughout the task (Bland & Schaefer, 2012). Unexpected uncertainty, on the other hand, can be understood as arising from prediction error that cannot be accounted for by current processing, exceeding the threshold of uncertainty tolerated by a chosen strategy, and making the cognitive system switch to the exploration of alternatives (Cohen et al., 2007).

In the brain, expected uncertainty drives tonic activity in the cholinergic system (acetylcholine, ACh) within the basal forebrain (including the hippocampus), which seems to optimize the activity of frontoparietal connections by modulating attentional episodes in the direction of task demands (Bucci et al., 1998; Aston-Jones & Cohen, 2005; Avery et al., 2012), and is able to maintain goals despite intervening distractions, a process that is also associated with tonic anterior cingulate cortex activity (Sadaghiani & D'Esposito, 2014). Unexpected uncertainty drives phasic activity within the noradrenergic system (locus coeruleus), with norepinephrine (NE) increases being able to reset prior probability distributions in order to develop new models of the environment, a process also associated with phasic activity in the anterior cingulate cortex and increases in midfrontal theta (Rushworth & Behrens, 2008; Cavanagh & Frank, 2014). Another proposed distinction is that of volatility, in which a prolonged period of unexpected uncertainty can derive into a model that takes into account that frequent uncertainty, producing either tonic NE increases (Bland & Schaefer, 2012), or tonic ACh increases (expected uncertainty) (Yu & Dayan, 2005).

Consistent with the hierarchically dominant role of the ventral stream, in Yu & Dayan's model (2005) the basal forebrain "gates the effectiveness of NE in controlling representational learning". In other words, prediction error may trigger decreases in ACh that are immediately followed by increases in NE, promoting not just a network reset but also plasticity mechanisms that can lead with time to consolidation and expected uncertainty.[9]

From this follows that the literature on exploitation and exploration is informative about the functional distinction that exists between the FPN and FTN, which can be summarized as the progressive allocation of resources towards bottom-up input as uncertainty increases. Based on that, the main prediction is that increases in ACh are

---

[9] Yu & Dayan (2005) also make the distinction between expected uncertainty due to the inherent stochasticity of the bottom-up input, or due to ignorance, with only the latter being able to reduce with time as more accurate models are established. This distinction matches with the proposal of the monophasic P600 introduced here, since this response seems to arise when persisting in solving a task in which the probability distributions of pre-activations are not yet set (i.e., ignorance). While the reduction of ignorance can be interpreted as the result of consolidation, it can also be seen from an online perspective as the accumulation of more evidence, which allows the model to become progressively more proactive. This is another way of understanding the progression observed by Chow and colleagues from the semantic P600 to the N400+P600 biphasic pattern when given more time to process role-reversed sentences.

related to dorsal stream engagement linearly, and that increases in NE reflect the coupling of the bilateral FPN with the dorsal stream. The simultaneous engagement of both networks may represent the possibility that both neurotransmitters, relying on partially independent systems, are not anticorrelated but can increase simultaneously. This simultaneous increase would combine unexpected uncertainty with the persistence necessary to match the unexpected stimuli with the ongoing representation (as opposed to abandoning the task set), and may therefore be informed by studies that specifically target the notion of volatility (e.g. Bland & Schaefer, 2012). There is also the possibility that NE increases phasically during the integration periods marked by the N400, signaling the flexibility necessary to move to the next integration.

In order to provide evidence for the continuity of both patterns of activity across the networks and its mediation by these neurotransmitters, an experiment can be conducted in which NE levels are boosted by a drug like idazoxan (Curet et al., 1987). The comprehension of long-distance dependencies in which working memory maintenance is assumed to take place is not always associated with increased FPN engagement, but this seems to be especially characteristic of older populations (Fedorenko, 2014). Peelle et al. (2010) show that for complex syntactic sentences older populations show compensatory activity in frontal regions outside of the FTN, less overall frontal activity, and less functional coherence. It would be interesting to see if performance and young-adult-like activity patterns can be recovered by administering idazoxan before processing these sentences. The reasoning is that comprehension in the absence of strong priors can be understood as a process in which unexpected uncertainty progressively turns into expected uncertainty and then certainty. The boost of NE could therefore speed up this process, leading faster or more efficiently to the proper levels of ACh to deal with the incoming input, and being reflected as increased FTN activity in an fMRI scanner.

1.4.5. A functional continuum of ERP and fMRI responses

The notion of uncertainty allows a characterization of the ERP and fMRI patterns behind the FPN-FTN relationship that stresses the idea of a single cognitive system shifting precision across different streams in order to efficiently reduce prediction error. The proposal is illustrated here by dividing uncertainty in two categories (expected and unexpected) with two degrees of intensity (low, high) that can increase in parallel for each category, and do so in a tonic or phasic manner. Although surely an oversimplification, this presentation helps us realize the continuous nature of the different patterns of activity and how they are related to each other according to predictive coding mechanisms, and introduces a way of exploring the responses from a domain-general perspective in which linguistic input can be just one medium among many others (with the FTN being substituted for any other domain-specific network depending on the task).

The suggested relationship is as follows:

Low EU/Low UU. Processing states in which there is almost no expected uncertainty (EU) can be characterized either by higher certainty or by less overall engagement with a task (less representational resolution, more distributed attention). The most certain states should probably occur on a subsecond basis whenever pre-activations are confirmed, and so they can be equated to attenuated N400 responses. As it will be discussed in the next chapter, higher certainty (low ACh and NE levels) sustained for a prolonged duration seems to imply a displacement of activity towards the ventral stream during processing, to medial regions, and a broader attentional scope associated with posterior cingulate cortex activity (Leech & Sharp, 2014). In sum, this is a state in which fewer metabolic resources are dedicated to processing, and can be associated in domain-general terms with the default mode network (Raichle, 2015).

Low EU/High UU. For states in which UU is higher and EU is still low, we may be talking about a lack of language proficiency, or the processing of less consolidated structures. Importantly, we can also include here those episodes in which the ventral stream has yet to respond or is uninformative about the bottom-up input, and therefore include conflicts arising from the lower-level dorsal stream instantiating monophasic P600 responses. As claimed before, this does not mean that the monophasic response has a unique nature compared with the N400+P600 pattern: in both cases there should be a shift of precision towards the lower levels of the rostrocaudal hierarchy, either because the ventral stream is uninformative or has mispredicted. In domain-general terms, we can associate this state with the more distributed, bilateral engagement of posterior regions with a less evident inhibitory influence from anterior regions.[10]

---

[10] The P600 has been categorized as an instance of the P3 component family (Donchin, 1981; Nieuwenhuis et al., 2005; Polich, 2007; Bornkessel-Schlesewsky et al., 2011), with the P3b response being associated with simple binary decisions more often than the P600, which may also be present for non-binary choices (e.g., during reference attribution), constituting a more inclusive index of the deployment of cognitive control mechanisms. Therefore, it seems tempting to ascribe the P3b to conditions of low UU, and the semantic P600 to high or more sustained activity related to UU. The P600 would be driven by increases in unexpected uncertainty that are more complex, perhaps requiring the multimodal combination of cognitive domains (cf. Brilmayer et al., 2017 for similar conclusions in the auditory domain).

High EU/Low UU. As EU increases and is tonically sustained, representational resolution increases, activity displaces towards left-hemispheric regions, and the ventral stream starts to process in a more proactive manner. The conditions in which the FTN engages are those in which low UU accompanies this activity. Significant monophasic N400 effects can also fit in here, signaling momentary increases in prediction error that are still tracked by ACh, or successfully contested (integrated) without the need to gather additional evidence. This association between the N400 and the FTN is often hinted at in accounts that describe the left hemisphere as the more "predictive" half of the brain, meaning the one most associated with the anticipation (pre-activation) of specific target words (hence, higher expected uncertainty) (Wlotko & Federmeier, 2007).

Still, the evidence supports interhemispheric cooperation as the cause of frontal negative responses (Wlotko & Federmeier, 2013), with the possibility that the right hemisphere plays a larger role than assumed in facilitating sustained FTN, left-hemispheric activity (cf. Poeppel, 2014). It should be noted that the link between the FTN and FTN + FPN patterns of activity is not equivalent to left-hemispheric vs bi-hemispheric processing, as at the former level of description we are specifically talking about where sustained activity occurs, whereas in the latter we are describing where processing takes place both phasically and tonically, and at multiple timescales, including those below the sensitivity of fMRI scanners. Therefore, the association of the FTN (left dorsal) state of activity with N400 effects is compatible with important right-hemispheric contributions to processing in the form of mostly uncorrelated activity. This will be explored further in the next chapter.

Additionally, the possibility of attaining progressively higher ACh levels in the absence of NE increases helps understand why the dorsal stream can activate with differing degrees of intensity without leading to the bilateral FTN-FPN pattern, for example during the proficient processing of long-distance dependencies (Gouvea et al., 2009). The reason why long-distance dependencies can also be associated with increased FPN activity (Peele et al., 2010) could be that there are actually two strategies to establish a connection between *filler* and *gap* (or head and dependency): by persisting with the same task set across intervening material (in which case an AP will be elicited signaling a proactive ventral stream that has found a match), or by searching through long-term memory once an unexpected input has been encountered (this idea will be revisited in Section 2.3).

In domain-general terms, this brain state can be associated with the engagement of BA 45, the apex of the rostrocaudal hierarchy (Nee & D'Esposito, 2016), which would exert an inhibitory influence over posterior regions, consistent with dynamic coding (Stokes et al., 2013, Hartwigsen et al., 2017).

High EU/High UU. During conditions in which EU is present, sudden surges of UU (phasic noradrenergic inputs) would signal large changes in the environment that violate prior expectations, or a transient decrease in performance and/or reward, engaging control mechanisms that suppress those priors and promote flexibility of thought (Avery et al., 2012, Bland & Schaefer, 2012). We may expect two potential outcomes. First, that at any time during this process of error detection a solution is found that involves persisting in the initial goal, leading to an N400+AP. This pattern can start within the N400 time window, but sometimes also within the P600 and between 500-900 ms (Federmeier et al., 2010; DeLong et al., 2014; Kutas et al., 2014). In this case, we should expect a distinct anterior increase of activity that is not present for the posterior positivities, signaling an early match with the ventral stream (e.g., hippocampal or prefrontal theta).

The second possibility is that a "network reset" (Avery et al., 2012) is triggered, abandoning the current interpretation in search for more evidence, thus triggering an N400 + P600 pattern. If this happens due to conflictive linguistic input, it should correspond with the engagement of the FPN simultaneously with the FTN, representing the generation of a new model based on the same bottom-up input, as opposed to a complete displacement of attention.[11] This may be captured by the notion of volatility, which may translate into neurochemical terms as a tonic increase of NE. Thus, a prediction is that this tonic increase of NE should be present for the posterior positivity only. In domain-general terms, this brain state would be associated with increased activity across bilateral FPN and FTN, and increased, sustained cross-network synchronization (Kitzbichler et al., 2009).

In sum, this perspective allows us to propose a continuum of responses, illustrated in Figure 1.1.

---

[11] The P600 has been recently related, like the P3, to a reorientation of the ventral attention network (Sassenhagen & Bornkessel-Schlesewsky, 2015), consistent with the idea of a NE increase causing a network reset. This finding also stresses the idea that ventral or limbic activity precedes and determines dorsal increases.

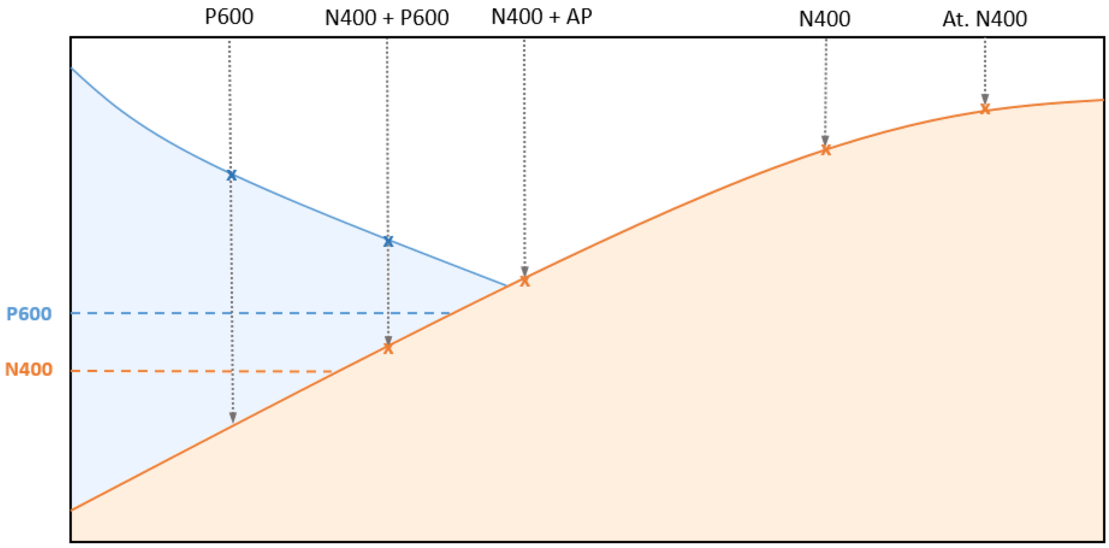


Figure 1.1. A continuum of event-related potentials across the dorsal and ventral streams. This figure illustrates a model in which each ERP pattern depends on the current state of both the dorsal and ventral streams, which are hierarchically related (with the ventral stream occupying a higher position). The vertical axis represents the activity of each stream, while the horizontal axis represents either time or degree of certainty.

When interpreting an input in the absence of strong priors, the dorsal stream (in blue) progressively decreases its activity as the ventral stream (in orange) takes over and explains away its prediction error, defining an anticorrelation in which the degree of certainty of the cognitive system increases linearly until a phase transition (when the two streams meet) takes place. This phase transition supposes that both streams synchronize and the ventral stream becomes proactive instead of reactive, i.e., capable of anticipating incoming input, and of overriding dorsal responses.

The horizontal dashed line in orange represents the threshold that needs to be reached by the ventral stream after which an N400 becomes likely, while the horizontal dashed line in blue represents the same for the P600 response and the dorsal stream. The latter threshold is crossed earlier in time than the former, as the dorsal stream is closer to the sensory signal than the ventral stream. This causes the monophasic P600 pattern.

As certainty and ventral activity increase further, and the ventral stream crosses the N400 threshold, there is a moment in time in which the N400 and P600 patterns are simultaneously found (blue and orange "x" symbols). If certainty can increase further (which is not always the case, as this depends on the input), the ventral stream takes over and posterior positivities turn into anterior ones, and then disappear, leading to an N400 effect that gets progressively more attenuated ("At.").

It should be noted that the graph can also be interpreted as the *result* of encountering a conflictive input; when this happens, the ERP patterns depend on how high ventral stream activity is at that moment, and how further "back" it is taken by the conflict before a repair is possible.

1.5. Summary and conclusions

The main goal of this chapter was to provide a framework for the relationship between the FPN and the two streams of the FTN based on predictive coding mechanisms. After dedicating some time to establishing a common terminology, introducing the phenomena to be explained and the predictive coding framework, this has been accomplished by suggesting a series of insights and research avenues that deserve further exploration.

Developmentally, the main claim was that the establishment of the adult version of the dorsal stream of the FTN should be considered as an ontogenetic process that departs from a domain-general, multimodal dorsal stream (the FPN), or a subset of it that is fine-tuned to auditory-to-motor transformations, which would make it language-selective from the start. This seems to be an alternative way of explaining language-specificity from a developmental as opposed to nativist perspective (Karmiloff-Smith, 2015), and it can account for the more bi-hemispheric and FPN-related activation patterns that are found in the infant or second-language learner brain, and for the evidence that the dorsal stream, the FPN and the DAN perform the same kind of operation. Also, this developmental perspective connects the use of the FTN with the processing of more consolidated, less flexible linguistic structures, in which the ventral stream plays a larger role in generating more specific pre-activations.

When characterizing the parsing mechanisms, multiple claims were made. The processing of non-adjacency was reconsidered as arising from an operation in which both streams have a necessary role to play. If a non-adjacent relationship needs to be processed, more prediction error is expected to arise within a dorsal stream that temporarily finds no ventral match that can explain it away as intervening material is processed. While this can still be understood as a working memory mechanism in action, it is no longer possible to characterize the FTN and the FPN as fulfilling computationally distinct functions. This was stressed in the following section (1.4), when it was suggested that both FTN and FTN+FPN activity can achieve the processing of the same structures, being only distinguished by the degree of certainty of the cognitive system.

Section 1.4 additionally examined how the ERP literature can inform these questions, and provided a picture in which parsing hypotheses that consider that the two streams are hierarchically related have an unexplored potential. Here, the main claim was that the phenomenon of pre-activation (the capacity to anticipate incoming input) is based on a set of graded or differentially weighted predictions whose success is indirectly reflected in the N400 response. “Ventrally constrained states” are those in which the specificity or weight of a candidate prediction increases over the others, either as a result of being dorsally processed for longer, or because, due to consolidation, a small amount of evidence is sufficient to induce it.

It was suggested, based on Chow and colleagues’ experiments, that the pre-activation of events in which arguments fulfill specific thematic roles is a costly process, as it involves both a high degree of specificity (constrains), and a high amount of information (complexity). Since the cognitive system strives towards efficiency (i.e., there is a tradeoff between metabolic costs and anticipatory power), the chosen strategy during reading or listening tasks is to use the linguistic signal to bias processing, as opposed to generating the event from a higher-level starting point.

This entails that the N400 is not necessarily insensitive to event representations, but that it will be more coupled to lexico-semantic access in these contexts. It also implies that the monophasic P600 for role-reversals is more likely to be found in contexts in which there is not a higher-level bias regarding event representations: in these contexts, the processing of the event representation occurs with the ventral stream taking a more reactive (instead of proactive) approach, with candidate events competing simultaneously with the processing of the linguistic signal (instead of a pre-activated event biasing it). In both cases, the ventral stream offers higher-level feedback to the dorsal stream, which forwards to it the prediction error that cannot be dorsally contested.

The switching of a strategy for another when subjects are given more time to process follows naturally from this model: more time spent without obtaining a reward should alter the values of the tradeoff, promoting the switch to a different strategy (exploration instead of exploitation). In this case, the new strategy is the simulation of the event without waiting for the linguistic signal, with the N400 + P600 pattern emerging as a consequence.

It should be noted that both monophasic and biphasic ERPs may only emerge when word order is the only source of evidence for determining thematic relationships. This makes sense if we consider that the suggested event representations that need to compete with each other in the absence of a strong event-like bias can only do so at the level of serial order, in which increased dorsal activity plays an established role. The hypothesis that event representations are universally costlier only works if languages that produce different responses from the monophasic P600 during role reversals do not provide a latency advantage when anticipating specific thematic structures. If that is not the case, it may be that events are only costly when depending on word order alone, as the cognitive system would be systematically used to not prioritizing other sources of information that can provide a latency advantage.

On the other hand, if events are universally costly to simulate, the N400 effect that occurs in some languages would need an explanation. It should be possible to look at increases of hippocampal theta preceding the target word to determine the extent to which the ventral stream is acting in a proactive (anticipatory) manner in languages in which an N400 effect is observed for role reversals in order to test the universality of this proposal. Indeed, the ERPs will probably be substituted in the near future by more precise oscillatory evidence when it comes to detailing how parsing works at the neural level.

Lastly, an important topic that has not been directly addressed is the contributions of the right hemisphere to the different patterns of activity across the FPN-FTN relationship. As an example, recent findings on the behavior of the inferior parietal lobe (FPN) during the processing of semantic anomalies seem to be in line with the model reported here, completing the picture. Tune et al. (2016) compare semantic anomalies that are easy to detect ("They were given the wrong meals by the *painter* and then they were overcharged for their meals") with borderline ones, which may even remain unnoticed by some subjects ("How many animals of each kind did *Moses* take on the ark?"), noting that the detection of any type of anomaly engages the left IPL, and the detection of easy anomalies specifically engages the right IPL. A possible interpretation is that right-hemispheric activity is associated with larger deviations from an established, highly weighted semantic context (i.e., *painter* is farther away from *waiter* than *Moses* is from *Noah*). This will be further explored in Section 2.2.2.

More experiments are needed to clearly establish the role of each hemisphere, but there seems to be a clear need for an “extended language network” (Hagoort, 2014) that takes into account right-hemispheric contributions that do not closely track the bottom-up input, but are nonetheless essential for comprehension.

In sum, the FPN-FTN relationship can be modelled on the assumption that the ventral and dorsal streams are hierarchically related through a predictive coding mechanism during all instances of processing, with FPN synchronizations signifying increases in prediction error triggered by uncertainty. This means that the switch from one pattern of activity to the other depends on the exploration-exploitation tradeoff, acting like a complex system which can be mechanistically described, but not fully predicted.

***

2. The ventral stream – a domain-general, hierarchical predictive coding perspective

This chapter focuses on the ventral stream of the frontotemporal network (FTN), asking how its activation patterns can be understood from a domain-general perspective, and how such an account can inform neurolinguistic theories. In particular, it explores the conditions that engage structures like anterior temporal lobe, inferior frontal gyrus, and ventromedial prefrontal cortex; and white matter pathways like the uncinate fasciculus and the extreme capsule fiber system, or the left and right ventral streams. This analysis is in turn anchored by a single linguistic question: is there a natural division between local and complex syntax at the brain level? The idea that non-adjacent linguistic structures are supported by dedicated brain mechanisms that are human-specific seems to be in direct confrontation with the domain-general, continuity-based perspective developed here. The main proposal of this chapter is that if the different activation phenomena of the ventral and dorsal streams can be understood from a domain-general, hierarchical perspective, such claim can be averted.

The chapter is organized as follows: first, there is a critical review of the local/complex division, including how it has evolved and changed locations in recent years, and what alternatives exist. There is a second section in which the idea of certainty is revisited to dispel the notion that only the dorsal stream implements predictive mechanisms, and that these are optional. This illustrates why exclusively structural notions like local and complex syntax, which do not take prediction into account, cannot effectively explain the observed patterns of brain activity, and also reinforces the view that dorsal-ventral interactions are necessary for all linguistic functions. The section continues with a discussion on the “left anterior temporal lobe effect” to examine the causes why the ventral stream is often exclusively associated with the processing of objects and local hierarchical relationships, and proposes a more nuanced perspective in which the ventral stream fulfills broader functions in concert with the dorsal stream. Lastly, there is a section on how to characterize complex syntax from a domain-general, functional connectivity perspective, which ends up stressing the granularity mismatch problem (Embick & Poeppel, 2004) that exists with this classification.

2.1. The local/complex syntax debate, and its functional consequences for the frontotemporal network

The neurocognitive distinction between local and complex syntax is defined by the observation that different brain regions seem to preferentially increase their activity for different properties of the linguistic signal. Specifically, the processing of local syntactic structures, in which dependencies do not exceed the boundaries of a single phrase or at most connect two phrases that are adjacent to each other, preferentially engages regions belonging to the ventral stream. On the other hand, any complexity increases in syntactic structures, for example the processing of center-embeddings or object-first relatives, preferentially trigger dorsal stream activity (Friederici, 2009, 2012). The purpose of this section is to critically review different ways of approaching this topic, which can illuminate our understanding of how the two streams of the FTN function and interact with one another.

One of the first models that committed to a specific brain implementation of these concepts is Friederici's (2009, 2012), based on previous work by herself and others on the functions of the dorsal and ventral streams. This model attempts to solve a puzzle that emerged when it was found that both local syntactic structures (Friederici et al., 2006) and sound-to-meaning mappings (Rauschecker & Tian, 2000; Hickok & Poeppel, 2007; Ueno et al., 2011; DeWitt & Rauschecker, 2012, 2013) can be independently ascribed to the ventral stream. These two seemingly different functions (their difference stressed by a clear-cut separation in theory between syntax and semantics, and language and speech) were attributed in this model to two different pathways within the ventral stream: the uncinate fasciculus, connecting orbitofrontal cortex to anterior temporal lobe (ATL) would be responsible for phrase structure building, whereas the extreme capsule fiber system, which runs parallel to the former but starting from BA 45 and BA 47, and going through posterior temporal cortex up to occipital cortex, would handle sound-to-meaning mappings (Friederici & Gierhan, 2013).

Thus, lower-level auditory information would travel from auditory cortex to frontal cortex along the ventral stream, incorporating local syntactic information reaching the frontal operculum and eventually BA 44, and semantic information reaching BA 45.

While BA 45 would retrieve and control lexico-semantic information from middle temporal gyrus (MTG), BA 44 would be responsible for building more complex syntactic structures and relaying them via the arcuate fasciculus to posterior temporal cortex (cf. also Hagoort & Indefrey, 2014), where both syntax and semantics are thought to be integrated.[12] Additionally, the superior longitudinal fasciculus, dorsally connecting posterior temporal cortex with premotor cortex, would be responsible for sensory-to-motor mappings and speech repetition (Hickok & Poeppel, 2004, 2007).

The purported independence of the notion of complex syntax was later supported by experiments that showed that the amount of activation, resting-state functional connectivity, and white matter integrity of the arcuate fasciculus positively correlate with the ability to process complex syntactic structures (Tyler et al., 2011; Vissiennon et al., 2016; Xiao et al., 2016). The protracted development of both the arcuate fasciculus and the ability to accurately process such structures, which progresses between the ages of 3 and 10 years, reinforces this link (Skeide et al., 2015; Skeide & Friederici, 2017; Fengler et al., 2016), as well as the much diminished presence of this pathway in the chimpanzee's brain (Friederici, 2017, but cf. Bornkessel-Schlesewsky et al., 2015b).

On the other hand, the link between the ventral stream and the processing of local syntax was weakened by studies that focused on the specific role of the left ATL for

---

[12] Another place of potential integration is inferior frontal cortex, as both BA 44 and BA 45 are highly interconnected, and respectively associated with the dorsal and ventral stream. For example, Hagoort (2005, 2013) proposes a *Unification Space* in inferior frontal cortex, where *syntactic frames* (local syntax stored in the lexicon) are retrieved from posterior temporal cortex and get assembled into a phrasal configuration for the whole utterance. Bornkessel-Schlesewsky & Schlesewsky (2013) propose a similar account for inferior frontal cortex, which they describe as a domain-general region responsible for both integrating the output of their largely independent dorsal and ventral streams, and resolving potential conflicts as a top-down influence of one pathway over the other.

linguistic combination. The initial association of the ventral stream with local syntax came from results indicating that damage to the left ATL leads to large impairments “at the most basic levels of constituent structure processing” (Dronkers et al., 2004), and that this region reliably activates for simple sentences, but only does it inconsistently for increases in syntactic complexity (Stowe et al., 1998; Friederici et al., 2006). More recent experiments like Brennan & Pylkkänen (2012; cf. also Brennan et al., 2016; Brennan & Pylkkänen, 2017) also show that ATL is the region most correlated with measures of constituent structure processing during natural comprehension conditions like listening to a story (cf. also Rogalsky & Hickok, 2008, but cf. also Nelson et al., 2017a).

Nonetheless, there is also a large body of evidence showing that ATL primarily responds to the presence of lexico-semantic information in the linguistic signal (e.g., Pallier et al., 2011; Matchin & Hickok, 2016), which is at odds with a purely syntactic conception of this region. Possibly as a result of these findings on the semantic sensitivity of the ventral stream, the latest instantiations of Friederici’s model have more clearly abandoned the idea that there is a purely syntactic dimension to this stream. Goucha et al. (2017) understand that the function of the uncinate fasciculus is limited to the processing of finite-state semantic associations, given that the frontal operculum is a phylogenetically older structure that engages during such conditions, when Broca’s region is relatively unresponsive (cf. also Friederici et al., 2006; Opitz & Friederici, 2007). This slightly modified perspective entails that both local and complex syntax would be associated with the dorsal stream, while the ventral stream would be exclusively semantic: “the ventral stream is rather dependent on associative learning based on transition probabilities in local dependencies and usually at fixed distances, allowing for the semantic combination of the sequence elements" (Goucha et al., 2017)".

The main account that questions the theoretical foundations of the previous model is Bornkessel-Schlesewsky & Schlesewsky (2013; Bornkessel-Schlesewsky et al., 2015a). These authors question why the ventral stream should handle such seemingly different cognitive functions as the processing of local syntactic operations and sound-to-meaning mappings, and why the local/complex divide, which there is no ground to separate in linguistic theory, should be handled by different streams. They propose an alternative model in which the two streams act in parallel to a large extent, with the ventral stream as a whole handling order-insensitive operations, and the dorsal stream being responsible for order-sensitive ones. This is the main functional division recognized by this account, which also denies the usefulness of the syntax/semantics division in neurocognitive theory. The advantage of this is twofold: it provides an elegant solution to the problem of how the ventral stream may be related to both auditory-to-meaning transformations and local syntactic operations: all order-insensitive aspects of both syntax and semantics are expected to appear within the ventral stream, which does not need to be sensitive to this cognitive boundary. Moreover, the distinction between local and complex syntax can be questioned on the grounds that the dorsal stream and inferior frontal gyrus (IFG) can engage for any order-sensitive (sequential) operation, including both local and complex syntactic operations.

The IFG is then seen as a mechanism for *linearization*, turning hierarchical inputs into serial strings for production (and doing the reverse for comprehension), while syntactic computations would actually take place in posterior regions (cf. also Boeckx et al., 2014). On the topic of language evolution, the authors understand that the arcuate fasciculus was enhanced in humans from a common substrate, triggering the possibility of acquiring and using more sophisticated forward models (sequential predictions) which could then be transformed as a top-down input for the ventral stream by prefrontal cortical regions (Bornkessel-Schlesewsky et al., 2015a). This evolutionary insight can also be applied to development, as the difference between local and complex syntax would be a gradual increase in sequential complexity reflected by increased engagement of the dorsal stream, and not a qualitatively distinct function implemented by different pathways.[13]

The debate is still ongoing. Newer instantiations of Friederici's model insist on the local/complex and syntax/semantics distinctions, and on posterior regions being responsible for integration. Their answers to the challenges posed by Bornkessel-Schlesewsky and colleagues seem to be that both the extreme capsule and the uncinate fasciculus are semantic finite-state processors, a macro-function that can encompass both auditory-to-meaning transformations and finite-state semantic structures. Regarding the criticism that there is no theoretical motivation to separate local from complex syntax, this is acknowledged by proposing that the relevant difference is between *Merge* and complex syntax. Merge (Chomsky, 1995/2014; Berwick et al., 2013) is the cognitive operation that takes any two items and combines them into a hierarchy in which one of the two elements (the head) projects towards the other. Merge builds phrase structures

---

[13] In this view, the higher-order function attributed to the ventral stream is the deployment of actor-event (AE) schemata, which form part of the order-insensitive, combinatorial operations seemingly instantiated by ATL. AE schemata are complex constructions with slots in which other constructions or concepts can fit, being thus integrated in an order-insensitive manner. They are abstract or semantic in nature, and not specified lexically for a particular word category, as the noun/verb distinction seems to emerge only after increased processing demands (cf. Vigliocco et al., 2011). In Friederici's model, the ventral stream may be too underpowered to handle AE schemata.

that are not limited by the finite-state capabilities of the ventral stream, and is also transmodal or content-blind, in the sense that it can build hierarchies across all domains of cognition, even when there is no clearly defined semantic content (e.g., in music).

In order to find the place of Merge in the brain, Friederici and colleagues (Zaccarella et al., 2015; Zaccarella & Friederici, 2015) set themselves to eliminate all confounding effects produced by semantics. To do so, they test two-word phrases with no lexical content ("this *flirk*"), which are then compared with word lists and single words. They find that the processing of minimal syntactic phrases engages the ventral-anterior portion of left BA 44. So the conclusion reached is that Merge, responsible for the local syntactic context, is located within BA 44, whereas complex syntax emerges gradually as the arcuate fasciculus, departing from BA 44, gains robustness across development, presumably allowing the formation of the brain-equivalent of more complex trees. In short, complex syntax amounts to the recursive repetition of Merge, a point that is stressed by the placement of both functions within the same stream.[14] As we will see in Section 2.2., these results have generated in turn a new round of challenges.

---

[14] Goucha et al. (2017) also defend that an additional operation is necessary to capture human-specificity: *Label*, which determines which of the two items will become the head based on category information. In their view, posterior temporal cortex integrates the outputs of Merge and Label via the arcuate fasciculus, yielding hierarchies that are interpretable by the interfaces. Thus, language reduces to the application of the ancient, associative, linear structures of the ventral stream to newly evolved, transmodal, dorsal hierarchies that exceed the computational complexity of finite-stateness, being capable of recursion. While both streams are necessary, the evolutionary novelty is dorsal in nature.

2.1.1. Main questions arising from the local/complex debate

We have seen multiple perspectives on what each of the two streams does, how they interact with one another, and how different aspects of linguistic structure emerge from these interactions. Some specific conclusions about what the main problems are can now be extracted.

First, there is the problem of how to characterize the function of the ventral stream, with some accounts downplaying its significance for language due to its responsiveness to finite-state sequences and associative semantics (Goucha et al., 2017), and others attributing to it a combinatorial character (Bornkessel-Schlesewsky & Schlesewsky, 2013; Brennan et al., 2016) that is very representative of what makes language human-specific. Some researchers would offer a compromise between both views by claiming that different linguistic operations rely on each stream (Bornkessel-Schlesewsky & Schlesewsky, 2013), or that both syntax and semantics have an independent combinatorial character in which there is no primacy of one over the other (Hagoort, 2013, cf. also Jackendoff, 2002, 2007). Following Chapter 1, the alternative view sponsored here is that of a hierarchical predictive coding architecture implemented within a single system, in which the dorsal stream is at the service of the ventral stream. Section 2.2.1 explores this relationship by claiming that the dorsal stream engages as a function of entropy, whereas the ventral stream does the same as a function of entropy reduction.

Second, there is the problem of what the posterior-to-anterior axis of the brain represents functionally, or what it means for IFG to engage. Some researchers claim that the role of IFG is to integrate the output of both streams and resolve conflicts arising from this integration (Bornkessel-Schlesewsky & Schlesewsky, 2013; Hagoort, 2013), with syntactic combination taking place in posterior regions; others claim that integration takes place in posterior regions, with transmodal hierarchies being formed within the dorsal stream, and acquiring their lexico-semantic information within posterior temporal cortex (Goucha et al., 2017). Since the current account assumes that the dorsal and ventral streams are in constant communication throughout all processing stages (with differing degrees of synchrony), that rostral and caudal regions act together as part of functional networks, that syntax and semantics are not natural kinds, and that the specificity and depth of comprehension varies with precision (cf. Sections 1.3, 1.4.2), no specific region for an integration function is proposed. Instead, integration should be understood as the cyclical achievement of a temporary equilibrium across widespread brain regions, similar to the global broadcast in the global workspace model (Baars, 1997; Dehaene et al., 1998). This account will therefore examine the specific roles played by ATL, BA 47, BA 45, and BA 44, and how they compare with posterior activity.

Third, the problem of whether there is a qualitative or quantitative difference between local and complex syntax seems to have been settled in favor of the latter view, as both Friederici and Bornkessel-Schlesewsky claim that complex syntax emerges gradually as a result of the dorsal stream doing what it does recurrently (even if they have different opinions on what the actual function is). Nonetheless, the debate is still far from being over, as this solution was already contested from a domain-general perspective in the previous chapter, where a necessary role for the ventral stream was advocated on the basis that complex syntax is related to conditions of expected uncertainty.

The present chapter also claims that the difference between local and complex syntax can be understood quantitatively, as just a matter of increased metabolic expenditure, a process that is affected by multiple contextual factors which are the real culprits of the emergence of one pattern of activity over another. Moreover, local and complex syntax would not be confined to any region in particular, but constitute general brain properties deriving from quantitative enhancements over a cingulo-opercular-frontoparietal-default-mode system. Section 2.3. characterizes this domain-general architecture, and demonstrates how the labels of local and complex syntax cannot adequately capture its complexity.

Lastly, there is the broader question of whether current models that focus on the FTN are able account for linguistic operations, or a domain-general perspective is justified. For example, language evolution is understood as a dorsal change because the assumption is that language can be characterized by a localizable, content-blind, unbounded operation for the formation of hierarchies (Merge) (Berwick et al., 2013; Chomsky, 1995/2014). This view should be contrasted with the observation that some combinations seem to be costlier than others because of their content, which dates back to the observation that center-embeddings are more difficult to process than what linguistic theory would suggest. In order to preserve the idea that language is truly free, part of the field reacted by attributing these problematic cases to other domains of inquiry: working memory, linearization, interfaces… The local/complex division seems to be another example of that strategy: while the basis for the distinction appears to be neural, based on whether the arcuate fasciculus engages more or less, the actual, underlying difference is better characterized as a cognitive one: working memory vs syntax (which there are reasons to believe are only distinguished by domain, not by mechanism, cf. the dorsal-FPN relationship defined in Chapter 1). This way of proceeding may not be useful if the intention is to connect language to phylogeny and the study of the brain, and domain-general alternatives are indeed necessary.

In sum, the next section focuses on establishing a single-system model that can account in a domain-general manner for increased anterior activity in each stream, and conditions of increased activity in the ventral over the dorsal stream. In order to succeed, it requires the abandonment of some conceptual distinctions (syntax/semantics, syntax/working memory), the adaptation of others (hierarchy/association, order sensitivity), and the inclusion of new ones that encompass a wider range of conditions (certainty/specification).

2.2. A predictive processing model of the ventral stream

Matchin et al. (2017) is a recent study that directly tests the validity of most of the functional claims described above, offering a new perspective on dorsal-ventral interactions. This section can be understood as a partial elaboration on this perspective, as it also argues against some of its assumptions. The ultimate goal is to provide a new, domain-general theoretical framework that can account for the observed patterns of activity without resorting to an opaque language-specific terminology.

The main question asked by Matchin and colleagues is why IFG and posterior superior temporal sulcus (pSTS), which are often taken as the source of syntactic combination, do not always increase their activity when processing syntactic structures vs word lists or baseline activity (e.g., Bemis & Pylkkänen, 2011; Friederici et al., 2010; Humphries et al., 2005, 2007; Rogalsky & Hickok, 2009). Their intuition is that the main function of the dorsal stream is not syntactic combination, but the generation of top-down predictions that are not necessary for language processing, but help by speeding it up, especially during difficult conditions like the processing of non-canonical structures. More specifically, the dorsal stream would be able to encode top-down predictions of phrasal nodes (NP, VP), syntactic features (agreement features like number or tense), and "perhaps syntactic information on specific lexical items or lemmas such as gender". Therefore, the authors claim that the fact that we sometimes observe significant dorsal stream activity may simply be the product of experimental conditions that encourage the deployment of such strategies, which otherwise are too metabolically costly to be worth it.

In order to test their "prediction hypothesis", the authors assume that two-word phrases (an example of local syntax) do not generally involve the use of top-down predictions, which are more likely to be used for non-adjacent relationships (complex syntax). In other words, they try to design a condition that is strictly bottom-up with the idea that it will not elicit increased dorsal activity, thus supporting their claim. While two-word phrases can elicit increased dorsal activity (e.g. Zaccarella & Friederici, 2015), they further attempt to minimize this possibility by letting subjects know that two-word phrases instead of full six-word sentences were going to appear, as they claim that this would prevent them from generating top-down predictions about a two-word phrase continuing and forming a six-word sentence. They contrast this condition with word lists, six-word sentences with simple structures, and the jabberwocky equivalents of phrases, sentences and lists, yielding a 3x2 design that allows them to investigate the specific functions carried out by multiple frontotemporal regions of interest.

Note that they do not directly test whether two-word phrases are predicted in a top-down manner or not. An alternative hypothesis acknowledged by the authors is that top-down predictions do exist for two-word phrases even in the experimental conditions they design; if that was the case, then the function of the dorsal stream would be the maintenance of such predictions across intervening words, which would capture why two-word phrases are not expected to elicit any increase of dorsal activity.

The main result of the study is that the two-word phrases that were tested do not seem to elicit significantly more dorsal or ventral activity when compared with word lists, both in the natural and jabberwocky conditions, a result that is at odds with Goucha et al.'s (2017) view that BA 44 is where Merge is localized in the brain. A potential explanation for why no contrast is found, according to the authors, is that Merge may be too fast to be reflected by the BOLD signal (a conclusion that fits with the view defended here that fMRI-observable activity mostly reflects prediction error deriving from the performance of faster operations). Additionally, sentences seem to elicit significantly more dorsal activity than phrases, but a similar amount compared to word lists. This is interpreted as BA 44 being involved in verbal working memory instead of syntactic combination, since two-word phrases presumably involve less storage costs than the other two conditions.

Seemingly not agreeing with these conclusions, a new review by Zaccarella et al. (2017) includes this paper but still defends that Merge is localized in BA 44. They specifically address the cases in which no dorsal activity is found for syntactic combination, and conclude that this is an artifact of using as a control condition word lists that contain both function and content words as opposed to one type only. This leads to the possibility of making partial combinations in word lists (e.g. “money the [the client washed]”, Kuperberg et al., 2000; “[Her eyes during close] the she ceremony”, Vandenberghe et al., 2002), thus yielding no contrast with the properly combinatorial conditions (i.e., leading to similar levels of BA 44 activation). When word lists are constructed with function or content words only, combinatory conditions reliably trigger BA 44 activity.

While Matchin and colleagues indeed use this kind of word lists, thus facing a potential confound, their results are not properly explained by such an account. This is so because the phrase condition seems to be unique in multiple aspects when compared with the other two conditions, despite the possibility of establishing mostly two-word phrases in lists. In particular, the phrase condition is not only processed with significantly less activity in BA 44 than the other two conditions, but the authors also report a trend towards significance in BA 47, BA 45 and ATL for natural phrases vs word lists, and in BA 47 for jabberwocky phrases vs word lists. While the first finding is interpreted as the result of reduced working memory costs, the latter two findings are tentatively suggested to reflect compositional semantics in the case of BA 45 and ATL (cf. also Pallier et al., 2011)[15], and of phrase structure building in BA 47. While the authors are not especially committed to the latter hypothesis, for Friederici’s model it would suppose the need to include the ventral stream as part of local, but truly syntactic operations (perhaps as an

---

[15] This interpretation can still make sense even if we acknowledge the presence of combination in natural word lists. The reason is that the kind of combinations allowed by word lists are very atypical semantically, abstract, or very difficult to imagine and understand. A quick look at the appendix that collects the stimuli of Matchin et al. reveals the following examples: “[close explorer] broker [the prayer phone]”, “[the nurse elf] our [the wind]”, or “[midwife spirit would] must [our should]”. As we will see, the interpretation defended in this chapter (that the ventral stream reflects certainty, and there is no natural syntax/semantics distinction) is also consistent with these results.

anterior source of very abstract labels). In the case of Bornkessel-Schlesewsky and colleagues, they can simply claim that this event represents the deployment of an AE schema, a conclusion that would resist a potential experiment in which no BA 47 activity is found, perhaps as a consequence of introducing non-mixed word lists as a control.

In sum, Matchin et al. (2017) is a very relevant study to disentangle all of the issues discussed so far. In particular, it introduces a framework to find out why dorsal stream activity seems to be contextually engaged, what relationship it holds with the ventral stream, and how this relationship changes depending on predictive mechanisms. The basic idea is that the dorsal stream instantiates optional top-down mechanisms needed for the processing of more complex structures or more difficult tasks, and that basic syntactic combination may not be localizable by fMRI experiments.

2.2.1. Increases in predictive certainty displace activity towards ventral regions

This section is dedicated to offering an alternative account for all of these findings based on hierarchical predictive coding, questioning the idea that prediction is optional and exclusively instantiated by the dorsal stream. The alternative offered here is that the dorsal stream engages as a function of expected and unexpected uncertainty, that activity displaces towards the ventral stream as certainty increases, and that the specifics of a task should determine the most efficient way in which both streams must engage to solve it.

The point of departure is the observation that prediction is not optional, but a graded phenomenon of varying efficiency, which in Section 1.4.4 was identified with the difference between the more reactive and proactive modes of operation of the ventral stream (in other words, how specific pre-activations are at a particular time, which depends on how confident the cognitive system can be about upcoming input). Once this is acknowledged, the different patterns of activity across the FTN can become informative about how well predictions are doing while processing a specific bottom-up input. More specifically, at least three conditions can be proposed: those that allow the cognitive system to have a high degree of confidence on their identity (high certainty), those that provide some obstacles that partially undermine this confidence (expected uncertainty), and those that either do not provide enough information to be interpreted with confidence, or that suddenly deliver an unexpected obstacle that cannot be surmounted by any prediction (unexpected uncertainty).

The proposal of the "prediction hypothesis" that exclusively bottom-up processing is not only possible, but the standard, is not compatible with a predictive processing perspective. Instead, this perspective invites us to consider dorsal engagement as a function of uncertainty, that is, how difficult it is to achieve a fully working prediction, a single interpretation among competing representations, a match between long-term memory and working memory... In sum, as a function of how much the current processing context manages to alter the homeostatic equilibrium of the cognitive system. Likewise, phasic increases in ventral stream activity can be understood as the bottom-up input being recognized, integrated, and in many cases fully anticipated; i.e., as a function of how much dorsal prediction error can be explained away by ventral predictions. Since the ventral stream is argued to be higher within the predictive coding hierarchy, this is often done at the expense of losing representational resolution, lower-level information, concreteness, etc. (for example, it is the reason why we can often remember the message of what someone said a while ago, but not their exact words; cf. Sanford & Graesser, 2006; Christiansen & Chater, 2016).

From this account, what Matchin et al. (2017) may be observing when they see dorsal stream activity would not be the optional engagement of top-down mechanisms, but an increase in their rate of failure (i.e., a partially uninformative ventral stream). This distinction is not simply terminological, because it entails that efficiency or speed are not simply a matter of which stream engages, but depend on the behavior of both streams and the specifics of the task.

Indeed, part of the evidence that Matchin and colleagues use to claim that the dorsal stream involves the optional generation of top-down predictions is that dorsal engagement is correlated with faster reaction times. For example, in Waters et al. (2003), the processing of object-first relatives was tested and participants were divided into two groups, one with faster subjects than the other. The faster group was found to have increased IFG activity, whereas the slower group showed left superior temporal lobe activity (both groups could successfully parse the sentences). More recent experiments allow us to be more specific: IFG in this case should refer to BA 45, the point of connection between both streams, which seems to preferentially engage for long-distance dependencies (Santi & Grodzinsky, 2007, 2010), but only provided that in these cases there is an active prediction waiting to be fulfilled (Matchin et al., 2014). In other words, BA 45 seems to engage for conditions of expected uncertainty in which a prediction (about a gap matching the filler) needs to be maintained despite conflicting evidence (intervening material) (cf. Section 1.3). Therefore, it should not be surprising that the best performing subjects in this task are those that can generate a prediction and maintain it, as opposed to those that do not do this and experience a larger surge of prediction error when the gap shows up. This means that the best performing subjects are adapting their cognitive system to the most efficient way of performing this task. However, if the characteristics of the task changed so that maintaining information against conflicts was not required, would reaction times still benefit from increased dorsal activity?

On the contrary, there is increasing evidence showing that dorsal engagement can also be detrimental for performance, as it can mean that a non-optimal *precision* (Feldman & Friston, 2010) is being employed when solving a task.[16] To put it differently, these are cases in which a dorsal increase would show that a prediction that should already be made is taking longer than expected. The causes can be multiple (cf. Chrysikou et al., 2014; Beaty et al., 2015, 2016), ranging from the cognitive system getting momentarily stuck in the exploitation of an ineffective strategy regardless of conflicting evidence; the lack of consolidated higher-level interpretations that can account for the raw sensory input, forcing an exploratory state; and the fact that some tasks can only be effectively solved by flexibly switching among representations (the *updating* function of working memory,

---

[16] Dopamine plays a major role as the mechanism controlling precision in the brain. The dopaminergic system is hypothesized to control the exploration/exploitation trade-off by means of the parallel stimulation of two different receptors, D1 and D2, with a D1-dominated state in PFC favoring robust online maintenance of information and increasing the signal-to-noise ratio of representations, and a D2-dominated state promoting flexible and fast switching among representational states (Durstewitz & Seamans, 2008; D'Esposito & Postle, 2015; Cools, 2015). The degree of incorporation of hippocampal-dependent memory in goal-directed planning has also been argued to be dependent on direct dopamine receptor stimulation in PFC (Goto & Grace, 2005; Cools & D'Esposito, 2011). The bottom-line of this literature is that adaptive behavior is not necessarily associated with persistent processing of information (frontoparietal activity), but with the dynamic adaptation of different cognitive states according to task demands (Cools, 2015; cf. also Cocchi et al., 2013; Chrysikou et al., 2014).

Miyake et al., 2000; or a high degree of *metastability* in network neuroscience, Kitzbichler et al., 2009; Shanahan, 2010).

One example of this in the language domain was already briefly introduced in Section 1.4.2: Ye & Zhou (2008, 2009) divided subjects into two groups, faster and slower, when processing subject-first and object-first sentences. Finding a violation of expectations during the processing of object-first sentences, which presumably engaged the mechanisms described before (BA 45 and dorsal stream), elicited a posterior positivity in both groups. However, violations in subject-first sentences, which do not require a maintenance or inhibition operation, triggered an anterior negativity in the faster group and a posterior positivity in the slower group, seemingly indicating that the faster participants were benefitting from more ventral stream activation when processing, and that they were processing proactively as a consequence, while the slower participants were relying on more dorsal activity.

Yet another example is found in Wlotko & Federmeier (2015), which measures the N400 response as a proxy of the ease of integration of novel information with the preceding context. They replicate the experiment in which subjects read a sentence presented in a word-by-word manner after being introduced to a context that was read in a self-paced manner, e.g. “They wanted to make the hotel look more like a tropical resort. So along the driveway they planted rows of PALMS/PINES/TULIPS” (Federmeier & Kutas, 1999). The second sentence is presented at 2 words per second (250 ms stimulus onset asynchrony) to one group, and at 4 words per second (500ms SOA) to another, and then this order is reversed. They find that at 250 ms SOA the reduction of the N400 caused by the semantic relatedness of PINES with the expected PALMS is diminished compared with a 500 ms SOA, demonstrating that pre-activations take time to engage. However, participants who read the stimuli with a 500 ms SOA first, and then read the 250 ms SOA block, continued to show the reduced N400 effect during the second block.

A conclusion we can extract from this study is that proactive predictive processing, associated with increased activity within the ventral stream and faster reaction times, engages as a function of certainty about how to process the stimuli, which can increase if subjects are already acquainted with the demands of a task. Again, there are reasons to think that the ventral stream instantiates predictive mechanisms as much as the dorsal stream does.

Results like these indicate that it is very likely that the two-word phrase condition tested by Matchin et al. (2017) is an instance in which the ventral stream is acting more proactively, given that the participants of the study were informed about the specific length of the structures before they appeared (i.e., certainty was made higher beforehand). This would indicate why two-word phrases would normally elicit significant dorsal activity (e.g., Zacarella & Friederici, 2015), but did not in this case: processing was biased to reduce the number of alternatives.

To find evidence for this, the pre-activated two-word phrase schemata of Matchin et al. (2017) should be contrasted with more uncertain conditions, and dorsal stream activation and reaction times should be measured in both cases. This has already been done in the domain of speech perception: Blank & Davis (2016) show that posterior superior temporal sulcus (pSTS), a posterior region that has a domain-specific preference for speech, engages more when speech information is less acoustically degraded (i.e., closer to natural stimuli). However, if there are prior expectations about the stimuli (e.g., subjects read a word right before they hear it), pSTS reacts in the opposite direction, with greater acoustic clarity leading to a reduction in the amount of speech information that can be decoded within this region by using multivariate fMRI. In other words, higher-level regions seem to be more easily explaining this information away.

The extrapolation of these results to our hierarchical model produces the following expectations. First, BA 44 would tend to increase its activity as a function of how much a specific input can potentially be processed, or how many alternative processing routes exist for it, provided that enough precision (top-down attention) is allocated to posterior regions. Second, if pre-activations are specific or constrained enough by the preceding context, the ventral stream will tend to engage more, with activity displacing towards higher-level regions (from ATL to BA 45) as a function of the divergence between the upcoming stimuli and such pre-activations (i.e., expected uncertainty). Third, the more effectively this difference can be reduced, the less activity will be found across the dorsal stream and the frontoparietal network (FPN). When the reduction is especially powerful (i.e., conditions of high predictive validity, in which pre-activations are fully specified and confirmed by the input), less activity will be found across the ventral stream itself. However, if the reduction fails, a large surge of dorsal activity will take place, involving the FPN.

An interpretation like this would allow us to understand the dorsal stream and BA 44 as part of a much broader network that is mediated by predictive coding, instead of as the location where working memory or Merge are contained. More in particular, this model can account for how dorsal activity is distributed across the three tested conditions (word lists, two-word phrases, and sentences): sentences and word lists elicit comparable

BA 44 activity, but two-word phrases elicit much less (cf. Figure 2.1), despite the facts that both sentences and phrases should be less demanding for working memory than word lists, and that both word lists and phrases possess very minimal syntactic structures. The reason for this would be that the sets of phrases, thanks in part to being pre-activated, entail a fast ventral integration that completely explains dorsal activity away at fixed time intervals (every two words), whereas sentences are larger and costlier event representations, and word lists contain structures of varying sizes (cf. Zaccarella et al., 2017) situated at unpredicted places (and in the natural condition, with uncommon meanings). The latter two therefore would imply an increase in alternative processing routes (an increase in the level of uncertainty) that is much reduced for two-word phrases.

Also, these results would show why the Jabberwocky conditions always elicit less dorsal changes compared with natural conditions (Pallier et al., 2011; Matchin et al., 2017), despite the unchanged presence of syntactic markers in them and their association with a semantics-free Merge (Zaccarella & Friederici, 2015). BA 44 would simply engage as a function of the amount of processing alternatives elicited by the stimuli, which are fewer when the content of words is unknown. Crucially, this is evidence that the dorsal stream seems to be affected by content as much as the ventral stream.

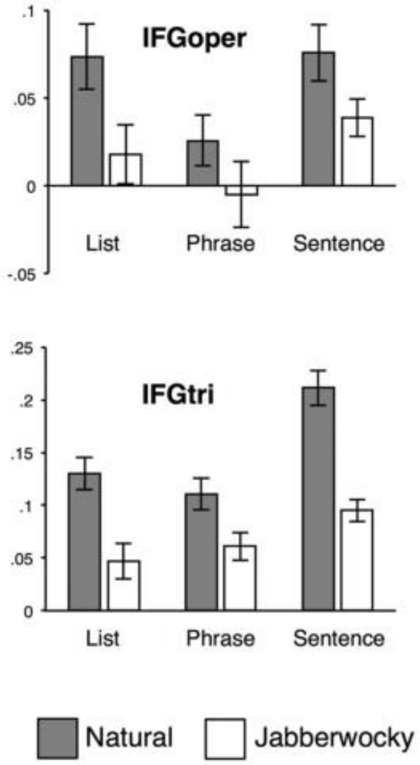


Figure 2.1. ROI percent signal change plots in BA 44 (*IFGoper*) and BA 45 (*IFGtri*) for word lists, three sets of two-word phrases, and six-word sentences (adapted from Matchin et al., 2017).

There is also a potential explanation for the different behaviors of BA 44 and BA 45: only when there is a complex event representation (i.e., in natural sentences) we find increased BA 45 sensitivity, whereas both lists and phrases elicit comparable changes (cf. Figure 2.1). If BA 45 engages as a function of the divergence between pre-activation and context within conditions of expected uncertainty, it should engage more when processing longer, meaningful sentences that by definition contain a richer, more detailed context that needs to be adapted. Meanwhile, BA 44 should also engage more for event representations that are more complex, as these imply more alternative processing routes (i.e., the uncertain part of expected uncertainty), but this is not the only reason it engages for: BA 44 is also triggered whenever BA 45 and the ventral stream are not acting efficiently at predicting the input (more unexpected uncertainty). Since both effects are mixed in the dorsal signal, we can argue that both word lists and sentences can elicit comparable BA 44 activity, but only meaningful sentences are associated with BA 45.

Lastly, the notion of certainty can eventually be formalized and tested with measures such as *entropy reduction* (Hale, 2016), an incremental complexity metric derived from information theory (Shannon, 1948), which determines how difficult the integration of upcoming input is given the immediately preceding context. Far from being limited to Markovian chains, this metric has been successfully applied to linguistic structure by Nelson et al. (2017b), using high-gamma power as measured by ECoG as the brain correlate of its theoretical predictions. In particular, the idea is that given an initial string there is usually a set of possible derivations that can be weighted, and more parsing difficulty (i.e., activation, high-gamma power) results from the cognitive system working to reduce the *entropy* (how equiprobable such alternative pathways are) of this set. Nelson et al. (2017b) show that the high gamma power of both anterior inferior temporal sulcus and posterior inferior temporal gyrus at about 400 ms correlate with word-by-word entropy reduction values derived from phrase structure grammars. This happens after controlling for factors related to word order and word identity by using word lists. In short, what this experiment suggests is that processing advances as uncertainty is reduced by the activity of the temporal lobe at specific points within the sentence that match the boundaries of phrase structures, and that its activity is inversely correlated with how efficient it can be at doing so (more equally weighted alternatives require a greater effort to disambiguate).

It should also be stated that naturalistic settings should offer more opportunities for efficient entropy reduction than most lab experiments, and that consolidation, short-term training or priming should also influence it by speeding up processing times. This may describe how the activity associated with parsing seems to be displaced towards the ventral stream whenever comprehension tasks are self-paced and offer rich contextual cues, and why in those cases brain activity is more closely followed by left-branching than right-branching parsers in experiments (e.g., Brennan & Pylkkänen, 2017 vs Nelson et al., 2017a). Most accounts would be forced to interpret this as the use of alternative processing pathways, top-down vs bottom-up strategies, or even as a semantic as opposed to syntactic way of parsing. An account like the present one, that considers that the BOLD response instantiates performance, and that behaviorally meaningful functions are the product of fast cross-network interactions, has an advantage for eventually explaining these results. In a nutshell, a reactive ventral stream should resemble more a right-branching, generative model of language, whereas an increasingly proactive ventral stream, by being able to reduce uncertainty earlier in time or more effectively, should resemble left-branching, construction grammar models of language, and involve less overall activity across the whole system.

In sum, the processing of phrases, sentences and word lists seems to be tapping on the effectiveness of the same dorsal-ventral interaction mechanism, which may be at the heart of not only syntactic combination, but of a common organizational principle of brain function based on predictive coding mechanisms. Having described how these mechanisms apply to the behavior of BA 44 and 45, we now move onto ATL, BA 47 and vmPFC in the following sections.

2.2.2. The domain-general object identification advantage of the left ventral stream

This section reviews and offers a domain-general perspective on the "left anterior temporal lobe effect" (LATL effect), one of the most studied behaviors of the ventral stream. The goals are to convey the idea that the ventral stream as a whole is much more powerful than what these findings might suggest, as its right-hemispheric and tonic activations, and its event sensitivity should also be taken into account; and to show how the notion of local syntax falls short of providing an account of this effect.

As advanced before, there was an initial period of research in which the ventral stream was considered to be as fully syntactic and language-specific as the dorsal stream. For example, Stowe et al. (1998) showed that only simple sentences activate bilateral anterior temporal lobe when compared to word lists, Dronkers et al. (2004) associated damage to this region with difficulties "at the most basic levels of constituent structure processing", and Humphries et al. (2005, 2007) distinguished a syntactic region in the most anterior portion of anterior superior temporal sulcus (aSTS). This trend has continued in recent times with studies like Brennan et al. (2012, 2016) and Bemis & Pylkkänen (2011, 2012), in which the temporal lobe is consistently associated with the processing of linguistic structures. Matchin et al. (2017) also found that BA 47, directly connected to the temporal pole, trended towards significance when processing local syntactic structures in both jabberwocky and natural conditions.

Nowadays, this perspective may have fallen out of fashion because of two sets of results: jabberwocky contrasts that show strong associations between natural conditions and the temporal lobe (e.g., Pallier et al., 2011), and studies on the specific semantic conditions that elicit left anterior temporal lobe activity (the “LATL effect”). The proposal of this section is not claiming for a return to a syntactic conception of this stream, but to offer a more nuanced perspective in which the syntax/semantics and dorsal/ventral divisions are not equivalent.

The anterior temporal lobe (ATL) is a hub that receives input from various sensory systems, directly connects to limbic and paralimbic regions (Patterson et al., 2007; Olson et al., 2007; Tune et al., 2016), and is generally involved in memory retrieval processes. It possesses a complex set of subdivisions that seem to organize around domain-specific and multimodal features (Binney et al., 2010; Skipper et al., 2011), and is traversed by the uncinate fasciculus in the temporal pole and the extreme capsule fiber system in anterior superior temporal sulcus. The LATL effect is a phasic increase of activity in left anterior temporal lobe that occurs at 200 ms post-stimulus onset, and is generally followed by activity in vmPFC at 400 ms (during language production, however, both vmPFC and LATL activate simultaneously at 400 ms; Pylkkänen et al., 2014).

The most recent models of the effect (Pylkkänen, 2017) propose a unified function for the ventral stream, *conceptual specification*, as a way of merging two parallel literatures: studies that show that the temporal lobe is sensitive to combinatorial semantics (Hickok & Poeppel, 2007; Bemis & Pylkkänen, 2011, 2012, 2013), and those that show that the LATL effect is involved in the online attribution of semantic features to entities (Baron et al., 2010; Baron & Osherson, 2011; Westerlund & Pylkkänen, 2014), with its damage leading to the loss of more specific concepts first (e.g. *trout* before *fish*) (Chan et al., 2001; Ueno et al., 2011). In short, conceptual specification is the notion that the LATL is generally involved in the online selection of specific aspects of one or more semantic association networks for processing, both within and across words.

Thus, Westerlund & Pylkkänen (2014) find that more specific words like *canoe*, when presented in isolation, cause a slightly larger amplitude increase in LATL when compared to more general ones like *boat*. Moreover, when these nouns are modified by adjectives like *blue*, the act of modifying *boat* produces a larger reaction in LATL than in the case of *canoe*. Zhang & Pylkkänen (2015) complement these results by showing that the more specific modifier *tomato soup* produces more activation than the vaguer *vegetable soup* at the time that the second word appears.

Also of note is the fact that LATL seems to have a preference for single entities to which different features are attributed, as opposed to sets. There is a LATL effect for sentences like “The girls are tall and blonde”, but not for “The girls are Finnish and Dutch”, if the second sentence is understood as some girls being Finnish and others Dutch (Poortman & Pylkkänen, 2016). In this case, it is right anterior temporal lobe (RATL) the structure that shows an effect, but at 400 ms. Perhaps similarly, “red cups” engages LATL in comparison to “two cups” (Pylkkänen et al., 2014).

It should be noted that these results are not explainable by traditional conceptions of combinatorial semantics or syntax in which operations are content-blind. Moreover, the effect happens very early on, at 200 ms after stimulus presentation, a time traditionally attributed to lexical access in middle temporal and posterior parietal regions (Van den Brink et al., 2001; Binder et al., 2009). When the LATL effect was thought to be syntactic (e.g., Bemis & Pylkkänen, 2011), this was convenient, as it was possible to suggest that syntactic processes came in first, initially treating all lexical items equally as atoms, and acted upon the interfaces at a later stage. Now we have evidence that one of the regions that first reacts is one that is simultaneously sensitive to lexical-semantic meaning and combination, and that is also situated within anterior regions, away from the posterior regions that supposedly handle integrations (but cf. Neufeld et al., 2016).

The LATL effect is still not fully captured by the models described in the previous section. Friederici and colleagues' model accounts for the LATL effect by proposing a finite-state parser in the ventral stream that is responsible for conceptual associations, but this is problematic when Merge is considered to be confined to BA 44, as an additional explanation is missing for why fully linguistic and combinatorial structures like *blue boat* can emerge without the need to rely on Merge at all, even though they are mostly not shared with other animals.

Bornkessel-Schlesewsky and colleagues' dual stream model also leaves some questions unanswered. Increases in ATL activity are taken as a measure of unification demands imposed on AE schemata, capturing the fact that this region can track constituent size increases, especially when semantic content is present: "unification occurs by incorporating one schema into a slot (e.g., 'who' or 'what') of another" (Bornkessel-Schlesewsky et al., 2013). Since they also claim that the ventral stream is responsible for the tracking of auditory objects of increasing complexity in an order-insensitive manner, they can also account for the observation that the LATL effect has a preference for single entities. However, the claim that lexical specification takes more effort and therefore a dorsal contribution is problematic in light of the lack of dorsal stream activity that is observed during LATL effect conditions; the roles of RATL and vmPFC are still undefined; and no account is offered on why the ventral stream is consistently engaged for these conditions even though the dorsal stream would be a valid alternative route for their processing.

The first thing to realize in order to tackle these challenges is that the LATL effect is not the ventral stream, but a specific, left-hemispheric activation pattern within it that arises under conditions of high certainty, when bottom-up input can be predicted in the most efficient manner. These cases were described before as resulting from subjects receiving instructions about the stimuli that they were going to encounter, which should have biased the ventral stream towards producing very constrained pre-activations. However, we may be dealing with a slightly different situation here, giving that the LATL effect is observed at 200 ms and is characterized by increased ATL activity, instead of the attenuated N400 that results from the combination of high certainty and predictive validity.

Following the previous chapter, the account provided here is that the cognitive system is structured so that it has a domain-general preference for the processing of objects, which can percolate to language in the form of a faster identification of highly consolidated linguistic structures referring to single entities. This link is stressed by the fact that, while the 200 ms time window is often associated in the neurolinguistic literature with lexical-semantic access and selection (Van den Brink et al., 2001), it is associated with attention and conscious discrimination in the visual literature (Patel & Azzam, 2005), suggesting a deeper link at the level of entity identification.

This proposal is inspired by the evidence that newborn infants are pre-equipped with *core knowledge systems* for the recognition and processing of (animate and inanimate) entities (Spelke, 1990; Spelke et al., 1992; Woodward, 1998, Sommerville et al., 2005; Kinzler & Spelke, 2007; Carey, 2009), which generate basic expectations about their physical properties and behaviors without the need for a costlier learning process. Core knowledge systems are suggestive of a very early and automatic organization of the raw sensory input that can bootstrap the development of more specialized cognitive contents and behaviors (Karmiloff-Smith, 2015), an insight that can also be extended past acquisition in order to characterize how online processing proceeds across stages.

Another source of evidence can be found in the working memory literature, which has settled on the consensus that the focus of attention is limited to a single, featurally complex entity at a time, which can be represented with the highest resolution, and gets displaced by other entities as the focus moves serially across time and space (Oberauer, 2013; D'Esposito & Postle, 2015).

In the psycholinguistics literature, the preference of the linguistic system for objects is already well-established, with nouns being acquired generally earlier than verbs, especially when they refer to imageable entities (Gentner, 1982; McDonough et al., 2011). Also, the early identification of actors and their increased salience over other elements of a sentence seems to be a cross-linguistic universal that precedes the understanding of sentence level meaning (Alday et al., 2014). In addition, many studies associate the processing or production of nouns generally with either anterior or posterior temporal lobe, and of verbs with the dorsal stream (e.g. Shapiro et al., 2005), but this distinction can collapse and become temporal for both categories under task conditions related to “semantic control” (Tyler et al., 2001; Vigliocco et al., 2011), which is consistent with the view that the domain-general single-object advantage extends to language.

Note that the claim that objects have a processing advantage is different from the claim that the ventral stream is singlehandedly responsible for their processing. In the domain of vision, object representations are thought to arise from an early collaboration of widespread dorsal and subcortical activity over the ventral pathway (cf. Kravitz et al., 2013), with the dorsal stream possessing a magnocellular cytoarchitecture that allows faster conduction speeds than the parvocellular ventral stream (Laycock et al., 2007). This fact can be associated with experimental evidence showing that early attentional mechanisms when processing a scene can be characterized by the obligatory existence of *tracking indexes* or *preconceptual objects* (Pylyshyn, 2001), which in turn drive attentional mechanisms in higher-order cortical regions (Cloutman, 2013).

Thus, the claim here is that there must be a role for the dorsal stream in the conditions that elicit the LATL effect, but limited to posterior regions, given the natural processing advantage that exists for consolidated objects. This entails that anterior dorsal activity should only accumulate under less consolidated conditions, like when exploring among a set of potential meanings for an unknown object, or when processing very novel combinations or non-literal meanings. In other words, the speed and capacities of the cognitive system to explain prediction error away and integrate anterior and posterior regions should depend on the content and specifics of the task, with the 200-ms time window being enough when processing a sensory signal that biases the interpretation towards a known semantic network, and the 400-ms time window being the standard timescale in which more widespread regions like those of the dorsal stream and the right hemisphere can normally be factored in, but only when they are needed.

This entails that there may be some fully linguistic structures (mainly those that refer to featurally complex objects) which can be supported, depending on efficiency considerations, within both timescales and across different patterns of cross-stream activity, blurring the lines between the syntax and semantics, and local and complex distinctions. In support of this, there is the experiment by Wlotko & Federmeier (2015) described in the previous section, in which short-term training produced a bias that speeded up the formation of pre-activations, causing the cognitive system to operate on the 200-ms time window when it would normally do so at 400 ms.

Even if a specific combination has never been encountered before and is not pre-activated, the use of the left hemisphere during the LATL effect conditions also suggests that the ventral stream is acting proactively and exploitatively, consistent with the idea that integrations are occurring under conditions of high certainty. To elaborate on this point further, it is worth mentioning an experiment by Wlotko & Federmeier (2013) on the different hemispheric contributions to the N400. They presented sentences ending with differing cloze probabilities alternating between the right visual field (left hemisphere) and the left visual field (right hemisphere). They observed that lateralized presentation to the right hemisphere requires a greater than 50% cloze probability in order for the N400 response to be attenuated, whereas the left hemisphere, which “has a tendency for more efficient top-down connections between frontal and temporal brain regions”, only needs 10-50% cloze probability to show the same effect. Thus, they suggest that both hemispheres provide unique contributions to language processing in parallel, with the N400 response being an average of both. In the case of the left hemisphere, it adopts more easily a proactive strategy that relies on language production (i.e., simulation) mechanisms, committing earlier to the appearance of specific semantic features in the bottom-up input.

From this it can be suggested that an experiment that tests the LATL effect conditions with a right-hemispheric presentation should probably find that it gets delayed to the 400-ms time window, and that the testing of less frequent combinations (*angry table*) may have even more chances to trigger dorsal stream or FPN activity than normally.

Lastly, bihemispheric contributions like the one observed in Poortman & Pylkkänen (2016) (RATL activates at 400 ms for "The girls are Finnish and Dutch") can also be investigated under the previous paradigm, in order to distinguish this effect from conditions that trigger dorsal stream activity. A hypothesis is that these conditions are related to the frontal positivity, which in the previous chapter was described as emerging from the immediate readjustment of a pre-activation, but that can also be studied from the perspective of the object-related advantage nearly missing to accurately process the input. In support of this, Wlotko & Federmeier (2007) show that the frontal positivity is not observed when the two hemispheres cannot cooperate.

If this framework is on the right track, a replication of Poortman & Pylkkänen (2016) while blocking hemispheric cooperation should lead either to a posterior positivity instead of a frontal one, or to a larger N400 effect. In these cases, the sentence "The girls are Finnish and Dutch" could be more likely to be interpreted as all the girls having a double nationality (an intersective reading, instead of the default collective reading in which some are Finnish and some are Dutch). This would provide evidence for a hierarchy of interpretations that would mirror a hierarchy of increasing complexity in the ventral stream: from single entities in LATL at at 200 ms, to bihemispheric ATL contributions that can handle collective readings at 400 ms, to intersective attributions of features to sets that may be associated with more anterior left ventral activity, perhaps involving the IFG at 400 and 600 ms.

In sum, the LATL effect reveals particular local syntactic conditions by cutting across the purely structural definition of local syntax, and it does so based on principles that seem to go beyond the domain specificity of language. Moreover, it highlights the existence of a ventral hierarchy that in turn influences the engagement of the hierarchically inferior dorsal stream. In the coming sections, we will explore this ventral hierarchy in more detail, and describe how its functions extend beyond the ones proposed after observing the LATL effect.

2.2.3. The required synergy between the extreme capsule fiber system and the arcuate fasciculus under conditions of expected uncertainty

The following sections are about the higher levels of the ventral hierarchy, as they investigate the roles played by the extreme capsule fiber system (ECFS) and the uncinate fasciculus (UF), and expand on the distinction between the LATL effect and the pattern of activity found by Matchin et al. (2017) for local syntactic conditions. Their main goals are to model their relationship from a predictive coding perspective, and to stress the previous claim that local syntactic conditions cannot be associated with a single pattern of activity within the FTN. Before starting, it is worth recalling that the UF bidirectionally connects orbitofrontal cortex, including BA 47, and the temporal pole, and the ECFS does the same between BA 45 and aSTS.

There are two recent studies that target the question of what roles are generally played across linguistic conditions by the two main white matter pathways of the ventral stream. A contrast between them will allow us to draw some general conclusions.

Hartwigsen et al. (2017) use for the first time dynamic causal modeling to replicate an experiment by Baumgaertner et al. (2002), which tested how the ventral stream reacted to differing integration demands. In particular, they contrast pre-activated endings ("The pilot flies the plane"), unexpected but semantically legal endings ("The pilot flies the kite"), semantically anomalous endings ("The pilot flies the book"), and pseudoword endings ("The pilot flies the *floop*"). The results reveal that the confirmation of a pre-activated ending elicits less overall activity in posterior regions (pSTS, pMTG) than all the other conditions, and that, as integration demands increase, there is a greater inhibitory influence from anterior regions to posterior regions: unexpected but semantically legal endings elicit activity in ATL (aSTS), whereas semantically anomalous endings increase the inhibitory influence of anterior inferior frontal gyrus (aIFG) over posterior regions.

Hartwigsen et al. (2017) interpret these inhibitory signals as reflecting the degree of semantic control that needs to be exerted for the successful integration of the target word with the preceding context, and associate them with the N400. They further speculate that the inhibition from aSTS to posterior regions would be conducted via the left middle longitudinal fasciculus (LMLF), whereas the inhibition from aIFG would be conducted either via the ECFS or the dorsal stream (arcuate fasciculus, AF). Finally, it is also noted that there is no direct correlation between the behavior of the ventral stream, which only seems to respond to integration demands, and reaction times. Since this correlation is indeed present for the AF, the authors argue for a functional distinction between *semantic control* and *executive control*, each depending on the ECFS/LMLF and AF, respectively.

Barredo et al. (2013, 2016) study how episodic memory retrieval affects the behavior of IFG, proposing that BA 47-UF is responsible for *controlled retrieval*, a notion that seems equivalent to semantic control in Hartwigsen et al. (2017) except for the chosen white matter pathway to represent it. They also contrast this notion with *post-retrieval control*, which according to them is supported by BA45-AF. The difference between both functions is the following: controlled retrieval controls competition for long-term memory access, whereas post-retrieval control operates on the output of the retrieved memories in order to adapt them to the current processing context. In other words, this is the difference between trying to remember something, and trying to decide among a set of remembered candidates to produce a correct response.

Barredo et al. (2013) establish these categories by exposing subjects to specific associations between items during a training phase, and varying the amount of exposure so that some associations have more strength than others, and therefore are easier to recall during the testing phase. The harder an association is to remember, the more that BA 47 seems to engage. Also, when they ask subjects to indicate whether a specific association has taken place before or not, they contrast new items with old items that appeared before but associated differently. This allows them to create a condition in which there is a conflict over response selection: an item seen previously with a different association generates a tendency to respond affirmatively, but the correct response is to suppress such tendency. This conflict is the one that is specifically associated with BA 45 and AF engagement.

Comparing these experiments and the present framework provides a more complete picture of ventral activity. A first thing to note is that the results of Hartwigsen et al. (2017) seem to be compatible with the existence of a predictive coding hierarchy across the ventral and dorsal streams. Their conditions involve the presence of the same highly specific pre-activation that gets either confirmed or disconfirmed to various degrees. They notice that activity in posterior regions (i.e., prediction error) increases with the integration demands of the target word, despite the fact that integration demands are also associated with increased inhibition from anterior to posterior regions (i.e., prediction). This pattern of more activation that is simultaneously subject to more anterior inhibition mirrors the experiment by Blank & Davis (2016), and it is likely that a multivariate analysis for the high predictive validity condition (“The pilot flies the plane”) would be able to decode less information specific to the word *plane* in posterior regions than when testing a more neutral sentence like “The boy likes the plane”.

Moreover, Hartwigsen et al. (2017) do not distinguish in their experiment BA 45 from BA 47, covering both with the label *aIFG*. It is likely that they are actually observing effects that are closer to BA 45, given that this is the purported apex of the prefrontal cortex hierarchy (Nee & D’Esposito, 2016), and that they specifically mention the ECFS. We will return to this topic in the next section with a proposal for why BA 47 and the UF are only mentioned by Barredo et al. (2013).

The AF in Hartwigsen et al. (2017) also seems to behave according to the expectations set by the present framework, which allows us to offer a different perspective on the notions of semantic and executive control proposed by these researchers. Hartwigsen and colleagues note that aIFG-AF is the pathway whose behavior contextually associates with the difficulty and latency increases characteristic of cognitive control demands, whereas integration demands are consistently associated with aIFG-ECFS. However, the authors also consider the possibility that the AF may play a contributing role for integration demands, especially for the semantically anomalous conditions ("The carpenter builds the *wine*"). They state that their results simply do not allow for a dissociation of the precise contribution of both pathways, but we can attempt to make one.

The contextual engagement of the AF for temporary episodes of more effort seems to follow from the idea that this stream, together with the FPN, is sensitive to cases of unexpected uncertainty. In support of this, Hartwigsen et al. (2017) list the regions associated with these cases: supplementary motor area, left or right supramarginal gyrus, and right aIFG, that is, regions belonging to the FPN.

The additional association established here between the AF and expected uncertainty gives us reasons to think that this is not all that there is to the AF, and the results of Barredo et al. (2013) seem to support it. In their task, there is a *necessary* engagement of BA 45-AF for cases of post-retrieval control, where multiple candidates are retrieved and one needs to be chosen over the others. Why is BA 45-AF engagement not as transparent for Hartwigsen and colleagues, and BA 45-ECFS seems to be the pathway that is necessarily engaged instead? Note that both experiments involve the suppression of a more automatic response, and are therefore expected to require some degree of semantic control.

One possibility already stated in previous sections is that the dorsal stream engages more as a function of the amount of alternative processing routes. All the conditions tested by Hartwigsen and colleagues ("The carpenter builds the... [table/tower/wine/*kirst*]") seem to be much more constrained or less ambiguous than those of Barredo et al. (2013). In Chapter 1, the processing of cases like *tower* (aSTS-pSTS via LMLF) was characterized by the anterior positivity, suggesting an immediate ventral repair of the pre-activation as soon as the target word shows up on the screen, and therefore no need for any further dorsal engagement. But the interesting condition is the semantically anomalous *wine* (aIFG-pSTS via ECFS), associated in Chapter 1 with the P600 and further dorsal engagement, or the cognitive system obligatorily performing additional operations before being able to make sense of the input.

The hypothesis for why AF engagement is not so evident for the *wine* condition in Hartwigsen et al. (2017) is that it provides fewer or costlier opportunities for being effectively processed (i.e., rewarded), and so the cognitive system dedicates less resources to it. This is based on the idea that the P600 is associated with *decision certainty* (Dröge et al., 2016), so that the amplitude is actually larger whenever the cognitive system actually knows what to do after an initial moment of confusion. It can be claimed that *wine* is a P600 condition in which there is less to do than, for example, when interpreting non-literal meanings or when solving the task from Barredo et al. (2013). Perhaps what the subjects normally do is to give up on a coherent interpretation, or to immediately substitute the verb *build* for *drink* via a phasic ventral engagement.

This hypothesis can be tested by an experiment in which a non-literal meaning ("The carpenter builds the *arguments*") is contrasted with the *wine* condition. Also, it can be proposed that asking subjects beforehand to try to imagine as vividly as possible the sentences they are going to hear may actually increase the precision dedicated to posterior regions, prompting them to spend more resources on trying to imagine how a carpenter could possibly build wine, and thus recovering a stronger P600 and dorsal engagement.

In sum, what these observations suggest is that processing during conditions of expected uncertainty is necessarily associated with a tight collaboration between tonic BA 45-ECFS and BA 45-AF activity (with the additional engagement of the FPN as tasks get more difficult or unexpected uncertainty increases).

This is consistent with the single-system, predictive coding view defended throughout the present dissertation, and seems to be supported by studies like Rolheiser et al. (2011). These researchers used diffusion tensor imaging to analyze dorsal and ventral contributions in a battery of language tasks involving both production and comprehension, and the domains of phonology, morphology, semantics, and syntax. They note that phonology (closer to lower-level information) correlates more with the AF, and semantics (closer to higher-level information) with the ECFS, while morphology and syntax show more mixed results, being alternatively associated with one or the other stream depending on task and performance differences. Also, they show how syntactic production is more associated with the AF, while comprehension displays more distributed patterns of activity between the AF and the ECFS. This is in accordance with the view that the sensory signal biases processing so that alternative processing routes are greatly reduced and integrations can happen more effectively, and so it clarifies the association between language production and FPN engagement (Fedorenko, 2014). Rolheiser and colleagues conclude: “Significantly, these data depart from the concept of a dorsal/ventral separation of function in all but the most clearly defined phonological and semantic tasks".

The next section completes the model with a discussion on the potential roles played by vmPFC and BA 47, and the different ways of operation of the uncinate fasciculus.

2.2.4. The medial route of the uncinate fasciculus – controlled retrieval beyond the level of single objects

This section proposes that the uncinate fasciculus is the ventral pathway that, when increasing its activity alone, is the most related to processing conditions involving high certainty. This implies that the tradeoff between processing speed and (spatial and temporal) representational resolution is more skewed towards the first as more activity accumulates within the UF.

A first thing to note when investigating these claims is that there are at least three main patterns of activity within the UF that may need to be distinguished, as they could involve different levels of representational resolution and processing latencies. The first one is associated with LATL-vmPFC activity, and includes the subcase of the LATL effect described in Section 2.2.2, which seems to be limited to the fast processing of multimodal, featurally complex single entities, and that was proposed to be able to involve their respective linguistic structures after consolidation. We shall call it the *medial* route. The second pattern of activity involves lateral orbitofrontal cortex regions like BA 47, and is the one that trended towards significance in Matchin et al. (2017) for their local syntactic structures, and the one introduced by Barredo et al. (2013, 2016) as responsible for controlled retrieval. We shall call it the *lateral* route, and note that most of the UF fibers terminate here rather than in medial orbitofrontal regions (Von Der Heide et al., 2013). Finally, the third one involves the frontal operculum (FOP), and was hypothesized to engage the UF alone for non-finite state, associative, local structures (Friederici et al., 2006; Friederici, 2012). We shall call it the *opercular* route. In what follows, we will contrast these subpathways and offer some possible functions for them.

The idea that the UF is associated with high certainty at the expense of lower representational resolution is best exemplified by the medial route. So far, we have only referred to the role of ATL as a multimodal hub for binding features to single entities, but this process exceeds the boundaries of this region during LATL effect conditions, seemingly constituting a whole-brain integration. Neufeld et al. (2016) is a recent study that attempts to investigate it in more detail by replicating with EEG the experimental setup of Bemis & Pylkkänen (2011) (*blue car/cup car/rnsh car/yerl car*). They notice that the three non-combinatorial conditions pattern together despite the fact that *cup car* also has semantic content (and the potential to act as a compound). Furthermore, they observe that the processing of *blue car* is characterized by amplitude changes at three stages: first, there is an early negativity at -50:100 ms, preceding the onset of the noun, followed by a centro-posterior, slightly leftward increased negativity corresponding to the 200 ms LATL effect found by Bemis & Pylkkänen (2011), and a broadly distributed increased negativity at 400 ms that they tentatively identify with the N400 and semantic integration.

These results stress that the medial route is consistently more engaged not just by simple item retrieval but by online conceptual specification or controlled retrieval (Westerlund & Pylkkänen, 2014), while the patterning of *cup car* with the pseudoword and non-word conditions, and the effects of anticipation at -50:100 ms show how the effect is related to conditions of high predictive certainty. It should be noted that the authors interpret this early negativity as an example of “predictive structure building”, which they also define as the deployment of “an empty structural position” to which lexical items are bound at 200 ms. While this may look at first as a syntax-first perspective, it can also be argued that what these two definitions are actually telling us that the EEG signal is the indissoluble product of both content and structure from the very beginning, as opposed to a content-blind operation. This is so because the early negativity seems to strictly depend on the adjective *blue* being fully accessible and interpretable.

Regarding the engagement of vmPFC at 400 ms, Neufeld et al. (2016) observe that it is actually larger for the combinatorial contexts compared with the non-combinatorial ones, which seems to go against the idea that the N400 is attenuated for predictable contexts. This is expected by the predictive coding framework, as it seems to be the same effect as the one discussed in Section 2.2.3 for the dorsal stream, which seems to specifically arise in experiments that contrast conditions of differing integration demands: whenever there are more operations to perform, overall activity should increase throughout the whole network, reflecting both increased prediction and prediction error (cf. also Blank et al., 2016 for evidence that this is the case across the FTN). On the other hand, whenever we are contrasting conditions that require the same operations but are differently pre-activated, overall activity should decrease with the specificity of pre-activations if these are confirmed.

A still open question is what the role of vmPFC could be, especially for LATL effect conditions in which the ATL would seem at first to be able to handle the integration of features by itself. While there are cases in the literature showing that the learning of visual objects does not rely on the UF (reviewed by Olson et al., 2015), other accounts make the vmPFC an integral part of this process. For example, Bzdok et al. (2015) propose that object processing in domain-general terms primarily relies on the synchronized activity of vmPFC and ventral posterior cingulate cortex (vPCC), supporting the idea that object processing can span throughout the whole processing cycle and involve the whole brain.

One possibility is that medial frontal structures like vmPFC and anterior cingulate cortex (ACC), which are adjacent, highly interconnected and directly innervated by the UF (Dick et al., 2014; Wang et al., 2015a), are capable of inhibiting dorsal activity under high certainty conditions like the LATL effect. Support for this can be found in other instances in which ventral activity is capable of producing “network resets” of this sort: the ventral attention network is known to be able to interrupt the activity of the dorsal attention network (Corbetta & Shulman, 2002; Dosenbach et al., 2008; Vossel et al., 2014), and the hippocampal-vmPFC route has been hypothesized to have preference over the dmPFC-caudate nuclei connections, which would be initially inhibited by the dmPFC (Dahmani & Bohbot, 2015). Also, the ACC forms part of the cingulo-opercular network, often involved with the thalamus in the control of cross-network interactivity (Kitzbichler et al., 2009; Danielmeier et al., 2011; Sestieri et al., 2014).

Thus, it can be proposed that for the LATL effect the ACC-PCC route would be consistently engaged throughout the whole task (which would yield no easily detectable contrast with brain imaging techniques), and that vmPFC activations at 400 ms would reflect the transmission of predictions to posterior, dorsal regions, globally broadcasting

(Dehaene et al, 1998) the results of the integration. This is also consistent with the role of phasic ACC firing in specifying whether there is a need to exert cognitive control or not (Shenhav et al., 2013; Cavanagh & Frank, 2014).

We can further speculate on how the oscillatory components of the N400 may reflect this process. The delta oscillation would be a promising candidate to support the inhibition of the dorsal stream by the ventral stream, as medial frontal delta has been associated with concentration, top-down attention, and the inhibition of motor responses, as well as with performance in semantic relatedness judgments (Fernández et al., 2012; Harmony, 2013). Regarding the latter, it seems significant that the integrity and connectivity strength of the UF have also been associated with performance in specifically this kind of semantic tasks (Harvey et al., 2013).

The connection between vmPFC, the N400 and delta is also supported by Steele et al. (2013), a study in which the N400 component is found to be divided into a theta and a delta oscillation. These researchers state that the power of theta increases when processing unpredicted information, thus serving as an index of the incoherence of new stimuli with the preceding context (just like the N400 itself). On the other hand, the delta oscillation appears slightly later but within the 400 ms time window, and becomes more powerful if pre-activated information is confirmed by new input. Also, a different delta oscillation that Steele et al. (2013) associate with the P600 does the same. Thus, the decision certainty that is attributed to the amplitude of the P600 (Dröge et al., 2016) can already be found in the second component of the N400, perhaps reflecting how a global broadcast takes place. It could be interesting to test whether the theta oscillation of the N400 is the only one displaced into the 200 ms time window during the LATL effect. This would stress how the mechanisms that instantiate integrations may be distinct from those that subsequently inhibit the rest of the brain.

Finally, these conclusions allow us to interpret that the recent study by Ding et al. (2016, cf. also Meyer et al., 2016), in which phrase structure building is associated with increased frontoparietal delta power at phrase boundaries, may be a reflection of conditions that are more effortful than during natural language settings, therefore requiring greater inhibitory power than usual, and entailing a more reactive ventral stream. This is supported by their stimuli being presented at an unnaturally slow rate, and purposely devoid of predictability and prosodic cues. This would mean that enhanced delta power would be just one consequence of phrase structure building under some conditions, instead of the mechanism that is singlehandedly responsible for it.

2.2.5. The lateral route of the uncinate fasciculus – controlled retrieval of higher-level task sets under conditions of higher certainty

Regarding the lateral route, the first question that arises is why Barredo et al. (2013) found increases of BA 47 activity as a result of modulating the difficulty of their task, while Hartwigsen et al. (2017) found that activity displaced from ATL towards aIFG, given that both tasks involved the access of a local association. The reason may be that the latter condition involved the act of switching a current interpretation (the pre-activated word) for another one (the actual word), involving post-retrieval control processes instantiated either by a phasic ventral integration (the medial route), or by the engagement of the ECFS and AF routes when certainty was lower or competition was higher. Barredo and colleagues' condition, on the other hand, lacked competition, was associated with higher certainty, and therefore involved the lateral route of the UF, which seems to engage as controlled retrieval demands increase.

However, this interpretation does not clarify why Barredo and colleagues found lateral instead of medial route activity. We will now briefly expand on this route with the end goal of capturing Barredo et al.'s (2013) findings and the trend towards significance in BA 47 that Matchin et al. (2017) observed for the processing of their two-word structures.

Studies on the UF do not seem to pay much attention to the PFC structures that engage as a consequence of the tasks that they employ, but this may be a crucial aspect when attributing functions to it. A general trend that can be extracted from the literature is that BA 47-UF engages during episodic and semantic memory conditions that are somehow more complex than the simple identification of single entities. For example, the UF has been involved in primates in the learning of object-location associations when these are rewarded (Bussey et al., 2002), which may be related to the association of orbitofrontal cortex (OFC) with reward (Von Der Heide et al., 2013). Also, Harvey et al. (2013) show how when two semantically related items (*pine tree* and *palm tree*) are maintained in working memory, and then another item (*pyramid*) is added and a decision must be made with respect to which one of the previous items it belongs to ("the pyramids and palm trees test", Howard & Patterson, 1992), the integrity and connectivity strength of the UF is most strongly associated with performance.

Results like these speak about the reason why Barredo et al. (2013) may have found lateral instead of medial route engagement in their experiment, as it involved the controlled retrieval of the episodic memory that associated two different entities. However, they are still not very clarifying as to why Matchin et al. (2017) may have found similar results for their two-word structures, given that half of them indeed referred to single entities. Provided that this result is actually significant, which may certainly not be the case, we can attempt a potential explanation.

Another general trend that can be drawn from the literature on the UF is that this white matter pathway is often defined as optional or more redundant than other ventral pathways. For example, it is not consistently engaged by many semantic tasks, including those related to the learning of visual objects (Olson et al., 2015), and its surgical removal leads to transitory retrieval deficits that revert after three months (Papagno et al., 2011; Von Der Heide et al., 2013).

The present hypothesis is that the lateral route is involved in the proactive retrieval from long-term memory of abstract relationships between elements (*task sets*) that are relied to BA 45 as potential candidates to be adapted to the ongoing processing context. This strategy would result from the ability to process such relationships in an initially reactive manner, and undergoing a period of short-term plasticity or long-term consolidation.[17]

The apparent redundancy of the UF is then explained because of two reasons: first, a reactive ventral stream (exploration) can already allow an effective task completion (cf.

---

[17] The association between consolidation and OFC, and the possibility of performing this task in a reactive manner as well may put under a new light the proposed functional specificity of human OFC, which has no equivalent in non-human primates (Semendeferi et al., 2001; Neubert et al., 2014). Like in the case of the human-specific subregion of BA 44, discussed in Section 2.2.6, it can be argued that no single region may be uniquely responsible for human-specific cognition, and that quantitative enhancements that cover the whole brain may have more explanatory power (cf. also Chapter 3).

Section 2.3); second, BA 45-ECFS engagement, associated with expected uncertainty and a more proactive ventral stream, can already attribute a single task set to an ongoing processing context, and control the AF based on it, but possibly more slowly, with a single candidate to choose from, and with a higher representational resolution. Additionally, it can be suggested that this single candidate may be relayed as a chunk by vmPFC to BA 45, which would be responsible for unpacking it, while the lateral route would be reserved for simultaneously suggesting a larger number of candidate task sets or more abstract ones (cf. Donoso et al., 2014, in which a maximum number of four simultaneous task sets is proposed).

This hypothesis is supported by work showing that OFC lesions do not prevent the acquisition of associations and the making of correct decisions during task performance, but are linked to slower learning rates (Wilson et al., 2014). These findings allow Wilson et al. (2014) to propose a model in which lateral OFC (including BA 47) is responsible for encoding abstract cognitive maps that label the contingencies of whole tasks beyond the level of perceptual details, which may be superficially similar across them. This is also in consonance with our proposed association between the UF and abstractness.

These conclusions allow us to hypothesize on the trend towards significance that Matchin et al. (2017) found in BA 47 for their pre-activated two-word structures, and why it is different from the LATL effect. First of all, their stimuli are somewhat more complex than those of Bemis & Pylkkänen (2011), as they are divided into two alternating types: “the fencer the baby their bill”, and “must record might abdicate should dominate”. On the other hand, this condition seems to be the easiest to process of all, as determined by the level of BA 44 engagement (cf. Figure 2.1), and according to the arguments of Section 2.2.1 on its link to high certainty.

It can then be proposed that the processing of this condition is guided by the simultaneous presence of two different abstract task sets encoded after subjects become aware that a sequence of two-word structures will show up on the screen: the one that corresponds to “the fencer the baby their bill”, and the one that corresponds to “must record might abdicate should dominate”. Thus, BA 45 would detect, based on its connection with the current processing context in the lower levels, when one of the two alternative task sets suggested by BA 47 becomes reliable, driving activity in more posterior regions like the AF (Domenech & Koechlin, 2015).

Furthermore, it may be the case that BA 47 is not only able to suggest, but also to explain away the activity of BA 45, which would be consistent with the results of Matchin and colleagues. This could be indirectly supported by the model of Nee & D’Esposito, in which BA 47 to BA 45 connections are shown to be inhibitory and excitatory, depending on whether segregated or integrated processing between both regions needs to be promoted (i.e., a greater influence of long-term memory vs working memory, a more proactive vs a more reactive ventral stream). Ideally, we would find that there is an enhanced inhibitory influence from BA 47 to BA 45 in a dynamic causal replication of Matchin et al. (2017).

A possible way to test this hypothesis would be to only test only one type of two-word sequence and see if BA 47 still trends towards significance. If it does, it can also be the case that is responsible for a single, highly abstract and complex task set such as “Process three groups of two, regardless of the identities of the members”. In that case, the prediction is that a contrast in which a single, pre-activated two-word structure shows up could make the BA 47 effect disappear.

2.2.6. The AF-ECFS route – a ventrodorsal hierarchy for the predictive processing of events and sequential order under conditions of expected uncertainty

Moving now to the opercular route of the UF, this subpathway has been involved in the processing of local, finite-state, associative combinations based on the literature that links the ATL with semantic content (Goucha et al., 2017) (cf. Section 2.1). This depiction seems strange given that unexpected violations in both finite-state structures and fully linguistic center embeddings with no semantic content (e.g., sound tone sequences) involve FOP engagement (Friederici et al., 2006), and that ATL-vmPFC seems to be a distinct pattern of activity that does not seem to co-occur with those cases.

It should also be taken into account that the opercular route of the UF is more controversial than the other two, as it seems to be indistinguishable from the inferior fronto-occipital fasciculus and the ECFS (cf. Dick et al., 2013, based on Ludwig & Klinger, 1956), and has been described as being based on “a rather crude assessment of anatomy” (Von Der Heide et al., 2013), given that the UF is arguably not connected with IFG. This actually suits our purposes, as the hypothesis developed here is that the opercular route forms part of the ECFS instead of the UF, forming part of the patterns of activity studied in Section 2.2.3, and of the rostrocaudal hierarchy of cognitive control.

Support for this can be found in a recent model by Wilson et al. (2017) that posits the existence of an evolutionarily conserved ventrodorsal prefrontal gradient for the processing of sequences of increasing complexity, which goes from adjacent relationships, to long-distance dependencies, to center-embeddings in cognitive terms, and from FOP/Anterior insula, to BA 44/45, to dorsolateral prefrontal cortex (BA 46/9) in neural terms (Wilson et al., 2017).

This conclusion is reached after observing that macaques and humans' brains are sensitive to violations in sequences with adjacent dependencies, which engaged homologous ventral regions of frontal cortex, including the FOP, and in the case of macaques including the homologue of IFG to a small extent (Wilson et al., 2015). Another comparative study (Wang et al., 2015b) also shows that when encountering violations in sequences of tones with respect to their numerical and sequential properties, both humans and monkeys engage comparable networks, but only humans show increased activity in bilateral IFG-pSTS for both types of violations, indicating that the human dorsal stream processes sequences indistinctively of their domain: "while monkeys can represent abstract properties such as 'four sounds' or 'one item is different', evolution granted humans with the specific ability to quickly unify these representations into a single nested structure such as 'three identical items, then a different one'" (Wang et al., 2015b). These conclusions remind us of Zaccarella & Friederici's (2015) view that a subregion of BA 44 is transmodal and a human-specific innovation.

Some observations regarding this literature are required. The first point to make is that the interpretation of bilateral engagement in both papers is different: Wilson et al. (2017) associate it with the participation of a more widespread domain-general network common to both species (Marslen-Wilson et al., 2014), while Wang et al. (2015b) see it as a result of using tones instead of language, that is, as a change of one domain-specific system for another. Given the association of the left hemisphere with more constrained pre-activations (cf. Section 2.2.2), and that these experiments do not require extensive training with the stimuli in either species, another (compatible) possibility is that these tasks initially reflect a lower certainty for both species than language processing would entail for humans. That is, both humans and monkeys start on more even grounds with respect to the task.

We can then propose that under these conditions processing efficiency increases as certainty increases, which directly depends on the contribution of the ventral stream (ECFS) in binding or integrating (or discarding, if these are not useful) patterns of dorsal stream activity that are initially parallel. In other words, there is a necessary ventral contribution for the learning and processing of this task, and the activation patterns observed in each species reflect the extent of such contribution. The manner in which it is reflected is the following: the more constrained ventral pre-activations can become during short-term learning, the more parallel dorsal stream activity can be reduced and confined to smaller regions, and the more upwards the ECFS ending in BA 45 will engage.

Let us now assume that the available metabolic resources or "cognitive bandwidth" of each species are quantitatively different. This entails two things from our predictive coding perspective: first, the lower the cognitive bandwidth is, the less effective the ECFS will be, which means increased dorsal stream engagement as prediction error spreads more freely (but only if the cognitive system still estimates that the effort of engaging with the task will be rewarded). Second, the higher the cognitive bandwidth is, the more opportunities arise for anticipating and explaining away dorsal stream engagement, which in turn frees up metabolic resources that can be dedicated to more dorsal engagement. In other words, there is an apparent paradox so that more dorsal engagement can mean both efficiency and inefficiency, or expected uncertainty and unexpected uncertainty.

This helps explain some of the observed patterns of activity across both species without assuming that single regions are responsible for single functions. The first one is the engagement of FOP for both local and complex syntax (Friederici et al., 2006), which can now be considered as an intermediate stage within the ECFS hierarchy, providing a standard anterior ventral contribution for the processing of conditions of expected uncertainty derived from sequential processing.[18]

As the representational requirements of the task increase, higher-level regions should be additionally recruited within the ECFS hierarchy terminating in BA 45, but only to the extent that expected uncertainty can still be maintained. If not, either the cognitive system will disengage with the task, will keep processing it at a lower level of complexity, or will attempt to engage with it by using a reactive ventral stream associated with unexpected uncertainty. The latter cases may help clarify why we find increased IFG activity in monkeys but not in humans for the processing of the same type of sequence of adjacent elements: the cognitive bandwidth of monkeys is lower, so they would start to adopt a more reactive strategy that nonetheless is still effective at solving the task (cf. output gating, Section 2.3). Alternatively, they would be able to dedicate less resources to it (a low expected uncertainty/low unexpected uncertainty condition), which would

---

[18] Just as event representations were depicted as costlier to simulate than objects (cf. Sections 1.4.3 and 2.2.2), the type of sequential processing in which the order of the elements matters must involve a higher spatial and temporal representational resolution than the processing of single objects, which entails more widespread prediction error, and the involvement of the ECFS and AF routes instead of UF. This is how the difference between order sensitivity and order insensitivity is understood in the present framework.

also be associated with dorsal rather than ventral IFG engagement (ideally, we would like to observe that IFG activity in these cases is more circumscribed to the monkey homologue of BA 44, rather than both BA 44 and BA 45).

Likewise, since FOP would be the ventral counterpart of BA 44, and the cognitive bandwidth of humans allows for a higher certainty, activity would rely on this more ventral region for the processing of the same structures.

This alternative interpretation would help solve an apparent paradox found in Wilson et al. (2017): they define IFG as a highly sophisticated sequential processor that has the ability "to extract and monitor for nonadjacent and multiple or hierarchical dependencies", but they observe that it activates in monkeys for very simple (adjacent) sequential tasks.

Regarding the experiment by Wang et al. (2015b), their association of BA 44 with a transmodal hierarchical processor seems to be an effect of the higher cognitive bandwidth of humans, who would have more advanced ventral integrative capacities, and would therefore be abler to reduce and integrate patterns of dorsal activity that are initially parallel and modular. In this case, the hypothesis is that BA 44 engagement reflects the uncertain part of expected uncertainty, entailing less overall activation across the whole brain, but increased activity in domain-specific, lateralized pathways like the AF and the ECFS.

These conclusions allow us to recast the finding of Zaccarella & Friederici (2015) that local, transmodal syntax (Merge) is associated with a subregion of BA 44. They test this under conditions that are not especially conductive to a proactive ventral stream, as they make an effort to reduce semantic content by using pseudowords. The activation pattern that they find seems to be a direct consequence of this condition: activity concentrates primarily within a subregion of BA 44. The location of this subregion in the *anterior-ventral* part of BA 44 (that is, in the direction of the FOP), and the apparent lack of involvement of BA 45, hints at the possibility that the ventral contribution for the processing of these structures is reactive (promoting BA 44 more than FOP) and does not dedicate as many metabolic resources (i.e., low expected uncertainty, low unexpected uncertainty) as it would during natural linguistic conditions, thus leading to the employment of a lower level of the ECFS hierarchy.

In sum, there are reasons to think that Merge may depend on cognitive bandwidth more than location, constituting a general brain property more than a novel operation. The last section of this chapter expands on this framework by studying the functional connectivity behind a domain-general cognitive system that can support all the patterns of activity discussed here within the context of the FTN, and argues against the separability of complex syntax based on it.

2.3. Complex syntax from a domain-general, functional connectivity perspective

This section argues that complex syntax is not special, but specialized. It is an attempt at showing that this phenomenon is not language- or even human-specific, but can be studied as a fundamental operation of cognition. The proposal to achieve this is to consider that the AF may be a distraction to fully understand complex syntax, and to focus instead on the fact that similar control functions in other domains are implemented by dynamic cross-network interactions that encompass widespread brain regions, so there is no reason why this should be any different for language. In this section, there is first a description of the domain-general functional networks supporting the working memory maintenance function, and then there is a proposal about how they connect with language. The section concludes with the idea that the differences between local and complex syntax are only quantitative, depending on the degree of metabolic expenditure.

Working memory maintenance is a domain-general label that seems to encompass the phenomenon of complex syntax. It refers to the idea that a cognitive item can be held in a buffer for a prolonged duration in order to associate it with another upcoming item, even in the face of distractions or the processing of intervening items. Maintenance is usually an effortful process, but this cost is offset by the more complex behaviors that it allows, which have been described as sequential hierarchies (Lashley, 1951; Fitch & Martins, 2014; Dehaene et al., 2015).

Given that there is an equivalence between maintenance and complex syntax at the cognitive level, it is surprising that the neural support for each concept is seemingly different. We have already discussed how complex syntax is currently approached: as the recurrent activation of BA 44 and the arcuate fasciculus (and also confronted it on the basis of what we know about the ventral stream and BA 45). What follows is a very brief history of the search for the brain basis of cognitive control, which until very recently was equivalent to the study of the maintenance function. We will argue that both language and cognitive control have followed parallel trajectories in the way they are studied, and that nowadays the state of the art is more advanced in the case of cognitive control if studied from a network neuroscience perspective. This offers some low-hanging fruits for neurolinguistics.

In the same way as how Broca's area became known as the "language region", cognitive control was initially confined to prefrontal cortex and defined as "the ability to orchestrate thought and action in accordance with internal goals" (Miller & Cohen, 2001). The study of functional networks modified this perspective in both fields, as soon linguistic processes were being described as resulting from collaborations between anterior and posterior regions within the FTN, and the same for cognitive control and the FPN (Cole & Schneider, 2007). Nonetheless, this new conception often kept implicit a locationist approach, just substituting the word *regions* for *networks*, which were still described modularly.

The following step was also a shared one: there was a recognition that other networks played a *peripheric* role for the functioning of the *core* network that supported each behavior (Bassett et al., 2013, Fedorenko & Thompson-Schill, 2014). In the case of language, we find the relationship with the FPN studied here in the first chapter. For cognitive control, some of the components of the FPN were recognized as clustering together as the *salience* or *cingulo-opercular* network (CON) (Seeley et al., 2007; Power et al., 2011).[19]

---

[19] Note that while subsequent papers have proposed to distinguish between the salience network and the cingulo-opercular network, situated slightly more posterior to the former (Power et al., 2011), this distinction is so far not routinely observed in the literature, which still keeps treating them as synonymous. Here, this distinction is recognized as the difference between phasic and tonic firing by the cingulo-opercular network.

It is interesting to note that the parallelisms do not only refer to the introduction of an additional network, but to the model of how they interact. We have already seen that the FPN normally acts independently but pairs with the FTN, which closely tracks linguistic stimuli, under contexts of increasing demands. The CON, comprising dorsal anterior cingulate cortex (dACC), anterior insula (AI), and thalamus, has two different ways of firing: tonic and phasic. Phasic activity is related to error and salience detection, and seems to specify the need to exert more top-down control (Menon & Udin, 2010; Collins & Koechlin, 2012; Cavanagh & Frank, 2014). It seems safe to assume that phasic CON activity is related to the N400 response, which is not necessarily followed by dorsal or frontoparietal responses, but can be under some contexts. On the other hand, tonic CON activity is always associated with the engagement of the FPN in the following way: as control demands increase, the CON and FPN start to synchronize with each other, with the FPN closely tracking stimuli (as in *dynamic coding*, cf. Section 1.1). In this case, the CON fulfills a function of *task set maintenance*, that is, it keeps the cognitive system engaged with the parameters and rules of the task while blocking distractions (Dosenbach et al., 2006; Cocchi et al., 2013; Sestieri et al., 2014; Sadaghiani & D'Esposito, 2014), similarly to how non-adjacent dependencies are thought to be processed.

This parallelism seems less mysterious when we consider that all networks are known to coalesce as control demands increase (Kitzbichler et al., 2011). For example, in an n-back task, in which subjects have to decide whether a stimulus matched another one presented a number of sequential steps before (Kirchner, 1958), performance seems to be correlated with the degree with which functional networks cluster together to form a global workspace architecture, in which information is exchanged freely across distant regions (Vatansever et al., 2015). It can be hypothesized that more effortful conditions in the linguistic domain not only trigger the pairing of the FPN with the dorsal stream, but also the integration of the CON with the FPN. Furthermore, the least effortful conditions of the n-back task are characterized by medial as opposed to lateral activations (Braun et al., 2015), similarly to the difference explored before between the LATL effect engaging vmPFC, and IFG engaging more generally. This allows us to establish a direct correspondence between the phasic way of firing of the ventral stream and the CON. In short, we may simply be describing how the same dorsal-ventral interactions that exist within the language network are also present in all other cognitive domains, with the CON playing a necessary role for all of them.

The centrality of the CON for all cognitive functions has already been discussed at length in the cognitive control literature. This is the point at which the neurolinguistics and cognitive control literature diverge, after the recent emergence of models that incorporate this network as a key component of what we know as control functions (e.g. Sridharan et al., 2008; Bressler & Menon, 2010; Menon & Uddin, 2010; Bonnelle et al., 2012; Chen et al., 2016). It is significant that this advance also entailed moving from considering that working memory maintenance and control were equivalent towards thinking about maintenance in terms of just one particular pattern of cross-network interactivity enabled by a broader notion of cognitive control (Kitzbichler et al., 2011; Betzel et al., 2016; Mohr et al., 2016). In contrast, neurolinguistics has been traditionally more restrained about the localization of its object of study, with the FPN and the ventral stream enjoying less overall attention than the arcuate fasciculus, and the CON being considered ancillary to language processes (but cf. Ferstl et al., 2008; Poeppel et al., 2012; Hagoort, 2014; Tremblay & Dick, 2016).

Emphasizing the importance of the CON for cognition and dorsal-ventral interactivity, Bressler & Menon (2010) propose that this network acts as an arbiter that determines, via its thalamic inhibitory connections, whether the brain will be dominated by a frontoparietal or default mode network state. These two networks are respectively thought to "undergo competitive interactions across task paradigms and stimulus modalities", and "mediate attention to the external and internal worlds". According to the authors, regions within the CON process sensory and limbic inputs, detecting salient events and initiating "appropriate control signals to regulate behavior via the ACC and homeostatic state via the mid and posterior insular cortex", thus promoting the activity of relevant networks and inhibiting irrelevant ones.

Completing another piece of the puzzle, more recent models of cognitive control also take into account how our conception of the default mode network (DMN) has evolved in recent times. This network comprises bilateral mPFC, MTL, posterior cingulate cortex (PCC), precuneus, and lateral parietal cortex, and was initially portrayed as increasing their activity during the resting condition, and decreasing it when subjects are asked to perform a task that requires external attention (Shulman et al., 1997; Raichle et al., 2001; Greicius et al., 2003; Raichle, 2015; Konishi et al., 2015).

Similar to the CON, the direction in which the conception of the DMN has evolved also shares some parallelisms with the portrayal of the ventral stream of the present dissertation. The idea that tonic dorsal stream activity requires a constant interaction with the ventral stream (cf. Section 2.2.3) is mirrored by evidence that the FPN and DMN collaborate when performing tasks. While transitory DMN activations are still associated with intrusive, unrelated thoughts and decreased performance (Weissman et al., 2006; McKiernan et al., 2006; Sonuga-Barke & Castellanos, 2007; Mason et al., 2007), its sustained engagement and correlated dynamics with the FPN have been shown to be beneficial for performance (Simons et al., 2008; Popa et al., 2009; Spreng & Grady, 2010; Gerlach et al., 2011; Smallwood et al., 2012; Fornito et al., 2012; Dixon et al., 2016).

Crucially, the condition that seems to apply to all of these tasks is a connection of online processing with memory, regardless of whether the task involves internally or externally oriented attention (Smallwood, 2013; Konishi et al., 2015). The same seems to be the case for the tonic engagement of the CON (Dubis et al., 2016). This allows us to propose that the dorsal cingulum (dPCC-dACC), engaging tonically and more anteriorly as integration demands increase, plays a unitary role in cognition that has much in common with the AF-ECFS within the FTN, providing a constant bottom-up-top-down exchange of information across the whole cortex.

The history of cognitive control does not end here. What has been described so far is a brain organized according to roughly static functional networks defined from the resting state. This depiction has been questioned by new models that take into account the fact that the same regions can participate in multiple networks (e.g., Van den Heuvel & Sporns, 2011; Leech & Sharp, 2014), and that cognitive processes are supported by whole-brain states (*meta-states*) with distinctive network architectures that evolve across short timescales (Chang & Glover, 2010; Cocchi et al., 2013; Hutchison et al., 2013; Zalesky et al., 2014). Ciric et al. (2017) is a recent model that attempts a synthesis of both viewpoints. The authors provide evidence for the existence of different network connectivity (NC) states within each of the canonical functional networks, and show that each NC-state can be associated to an extent with a distinct meta-state or whole-brain context: “information about the state of the entire brain is embedded in each network” (cf. also Braga et al., 2013).

Unlike more static descriptions, NC-states enable distinct regimes of dynamic cross-network interactivity. While NC-states had gone mostly unnoticed due the employment of time-averaged connectivity techniques, they can replicate common findings in fMRI experiments, like the merging of the CON with the FPN. They seem to be shared by all individuals, and constitute a challenge for the idea that the brain is organized by independent networks. However, Ciric et al. (2017) also demonstrate that when nodes are chosen randomly from the whole brain, creating *pseudonetworks*, these can inform us more accurately about whole-brain connectivity than NC-states, meaning that the canonical functional networks retain a certain degree of modularity and idiosyncratic within-network connectivity regardless of constant cross-network interactions. In fact, Ciric and colleagues argue that it is even possible that two networks exhibiting two concordant NC-states may have at some points not much connectivity between them.

In sum, we can conclude that the current implementation of cognitive control operations cannot be explained without appealing to rapid cross-network interactions that constitute a series of NC-states across different functional networks, and that in turn entail a specific whole-brain connectivity. This overview therefore takes us to the important questions of how to characterize complex syntax from this perspective, and of whether there is actually a one-to-one correspondence between complex syntax and a specific series of NC-states. The remainder of this section is dedicated to providing an experimental setup and some notes to eventually accomplish a full characterization, and also draws some conclusions about the local/complex classification.

There are two main proposals: that the NC-states characterizing working memory maintenance and complex syntax are equivalent, and that these are actually contained within the set of NC-states characterizing all instances of local syntax.

Support for these claims could be found by adapting experiments like Unger et al. (2016), an analysis of the dynamics behind the age-related performance enhancement in working memory tasks that is experienced during adolescence (Amso et al., 2014). In this experiment, a digit is introduced followed by a symbol and a letter in one condition, whereas in another condition the digit follows a symbol and a letter. In both cases, the digit determines whether the symbol and/or the letter are relevant in selecting a specific response, constituting a higher-order rule. This way, in the first condition participants (both children and adolescents) can implement the rule immediately and proactively, discarding irrelevant information as it appears on the screen (e.g., the symbol), whereas in the second condition participants have to wait until the digit appears, keeping both the symbol and the letter in working memory, and having to deal with competing rules reactively, at the moment that the digit shows up. These two conditions are respectively known as *input gating* and *output gating* problems. Importantly, the first condition can also be solved reactively (as an output gating problem) by loading everything into working memory and then deciding which rule needs to be discarded based on the digit that appeared first, but this strategy is less efficient, involving longer reaction times and more effort (presumably, more FPN engagement, although this was not measured).

According to the results, children and adolescents can maintain a similar number of competing associations in working memory, children prefer output gating strategies, and children perform worse with output gating strategies than teenagers despite their preference for them. In other words, maturation entails both an enhancement of output gating mechanisms, and an increase in the use of input gating strategies, while maintenance capacity is held constant. Note that input gating mechanisms did not seem to improve: when children used a proactive strategy, their performance was similar to adolescents'. The interpretation of the results offered by Unger and colleagues is that output gating mechanisms are handled by corticostriatal loops, which take time and experience to mature, slowly increasing performance in trial-to-trial control tasks (Frank & Badre, 2011; Chatham et al., 2014; Chatham & Badre, 2015). Regarding children's failure to engage input gating mechanisms, the authors suggest that it may be due to their being costlier to deploy than for adolescents. Furthermore, they associate these mechanisms with the implementation and sustenance of task instructions, whose increased cost they attribute to the less mature default mode network of children (Fair et al., 2008).

This experiment constitutes a useful paradigm because it specifically targets the different steps of working memory tasks (encoding, maintenance and retrieval) under higher-order conditions that are especially effortful, and establishes a contrast between different age groups and between a more proactive and reactive ventral stream. Adapting this experiment to language and contrasting the results would allow us to target empirically many of the claims discussed throughout this chapter. Just like the digit in Unger et al. (2016), a filler allows the proactive generation (pre-activation) of features belonging to the gap, which needs to compete against intervening material as processing advances. However, they can also be solved reactively (and less efficiently) when the gap is encountered, by performing a serial search through short-term memory until the filler is encountered.

It should be possible to replicate Unger et al.'s (2016) results very straightforwardly with a longitudinal study or by comparing different age groups. In particular, the expectation is that younger populations who can already process these constructions accurately tend to perform output gating operations for fillers more often than adolescents, and that adolescents can complete them faster than children. Brain imagining should reveal that the maturational state of both the AF and corticostriatal loops is correlated with less frequent and faster output gating operations, consistent with the strong association between both systems demonstrated by Jeon et al. (2014).

Crucially, the degree of FPN activation should be *negatively* correlated with the efficiency of output gating strategies, but *positively* correlated with choosing them over input gating strategies. It could even be possible that output gating strategies are completed via more FPN engagement by children, and more AF engagement by adolescents (in which case it would be difficult to tell the employment of an input over an output gating strategy apart, perhaps having to rely on reaction times). This would provide evidence for the continuous relationship between the FPN and FTN discussed in the first chapter. It would also show that complex syntax is a broader concept than the arcuate fasciculus: presumably, long-distance dependencies are acquired very inefficiently at first, relying on widespread frontoparietal and corticostriatal connections that are already much more capable than those of the chimp's brain. Due to consolidation, the possibility of turning them into input gating problems (i.e., to simulate them, to process them proactively) would increase, and even when failing to do so their reactive processing would be a much more constrained problem than before, entailing less dorsal activation (i.e., a confinement to left dorsal activity, associated with expected uncertainty).

From this it follows that the different activation patterns shown by working memory tasks and complex syntax are only different to the extent that they are the same mechanisms acting over different domains of cognition with different degrees of consolidation. In other words, the “arcuate fasciculus advantage” of the human species may only exist as a result of a quantitative enhancement (both in development and evolution) of the domain-general CON-FPN-DMN architecture, in which the vast majority of white matter pathways, which also have a comparable period of maturation (Asato et al., 2010), should play a role.

In indirect support of the AF being an effect of consolidation, and not the enabler of complex syntax, there is already work showing how extensive training with motor tasks like finger tapping sequences entails a similar disengagement of the CON-FPN with a dorsal stream, but in the right hemisphere (Bassett et al., 2011; Stout et al., 2015; Sale et al., 2017). The fact that both the within-network (NC-states) and cross-network functional connectivity of this dorsal subset change with consolidation (Cocchi et al., 2015), and that different connectivity patterns can emerge from the processing of the same structures, allows us to realize that the concept of complex syntax may not be adequate to describe brain activity beyond a very superficial level, and that NC-states may constitute a more accurate interface between cognitive description and the neural substrate.

The adaptation of Unger et al. (2016) to language can also be used some day to investigate the claim that the NC-states characterizing complex syntax constitute a subset of all the NC-states characterizing local syntax. There are two arguments for this being the case: that it is possible to devise local contexts associated with patterns of activity attributed to complex syntax, but there are some complex syntactic structures that may not be solvable by exhibiting patterns of activity characteristic of local syntax; and that all syntactic operations should require a dorsal-ventral interaction profile involving the brain as a whole. Once a set of NC-states emerges for input and output gating strategies across different ages, they can be searched for in local processing contexts that contrast different degrees of certainty (i.e., more vs less consolidated constructions, slower vs faster processors...).

This section will finish with a proposal on what to expect when trying to test long-distance dependencies from the perspective of NC-states in an experimental setup similar to that of Unger et al. (2016). The objective is to characterize the sets of NC-states that should be responsible for a filler-gap dependency as processed through an input gating strategy, and to distinguish them from other whole-brain regimes of information exchange.

For an input gating strategy, an initial dorsal-ventral interaction should instantiate the pre-activation to be maintained in working memory. According to what we know about the processing of working memory tasks, left hemispheric FPN and DMN interactions take place during the encoding phase, and the FPN and DMN stay separate during the maintenance phase (Piccoli et al., 2015). However, given that the processing of intervening words does not stop when a pre-activation exists, it is likely that we encounter recurring patterns of segregation and integration between both networks as processing goes on, and that we can only distinguish that a pre-activation is being maintained by other measures like an enhancement of the BOLD response or theta/gamma power, or small frequency differences within the theta-gamma code of working memory (Lisman & Jensen, 2013; Bahramisharif et al., 2017). It is also possible that the difference between input gating and missing to input gate only becomes noticeable at the retrieval phase (i.e., when the gap shows up): unlike in the encoding phase, the retrieval phase of working memory tasks is characterized by bihemispheric FPN-DMN integration (Piccoli et al., 2015).

In any case, it seems that an equivalent NC-state within the dorsal stream can instantiate both the maintenance of pre-activations and a lack of them. We can identify such NC-state with *FPN2* (*EXE2* in Ciric and colleagues' terminology), which occurs when the FPN nodes become segregated into two different subsets, with one exhibiting uncorrelated, cross-network activity, and the other modularizing. Ciric et al. (2017) claim that the latter subset may represent "a shift towards specific, localized computation in these subsystems", which is consistent with the idea that both the maintenance of pre-activations and a lack of them can exist within conditions of expected uncertainty.

Even if FPN2 was common to both pre-activating and failing to pre-activate, each case could still represent a distinct meta-state, as we know that there is an increased engagement of BA 45 when pre-activations are waiting for a match. Thus, the pre-activation case may correspond with greater anterior ventral activity, which might be reflected by a greater coupling of the CON with the FPN. According to Ciric et al. (2017), the NC-state of the CON that shows the greatest correlation with FPN2 is *CON2* (also known as *SAL2*), with *CON5* being the other possible candidate. There is some evidence to prefer CON5 over CON2 based on their relationship with the DMN, as we will see below.

The NC-state of the DMN that seems to emerge as a function of working memory load is *DMN4*, which Ciric and colleagues associate with “increased allegiance among salience, executive, and subcortical systems”. This NC-state takes place together with FPN2 and CON2 when cognitive capacities are working to their fullest, so this regime may be a good candidate to support center-embeddings and even the learning of filler-gap dependencies, but not the employment of an input gating strategy for processing the latter.[20]

This takes us to two better candidates: *DMN2* and *DMN3*. Ciric et al. (2017) describe these as showing a clear correspondence with the MTL and dmPFC subsystems of the DMN (Andrews-Hanna et al., 2010; Raichle, 2015), which temporarily dominate the integrity of the network. While each of them seems to engage independently, both share a high degree of connectivity with a third component of the DMN, the *midline core* composed of precuneus, PCC, right inferior parietal lobe, and mPFC. Ciric et al. (2017) show that when DMN2 is engaged, the nodes of the dmPFC subsystem demodularize, increasing their connectivity with other networks like the CON, and when DMN3 is

---

[20] A developmental account of how consolidated non-adjacent predictions develop that does not need the presence of high levels of uncertainty, but can start from very low-level prediction error within the FPN, can be found in the work of Onnis et al. (2003, cf. also Misyak et al. 2009; Onnis et al., 2015). In a nutshell, there is evidence that the learnability of non-adjacencies depends on the distributional properties of the input, so that both very high variability or zero variability in the words between non-adjacent dependencies facilitate the learning of non-adjacency by infants and adults, but a medium level of variability impairs it, since it contrasts less with the variability of the words that constitute the dependency. This would be an example of how the development of complex syntactic structures would not necessarily rely on the development of a robust arcuate fasciculus.

engaged, the same happens to the MTL subnetwork, increasing its connectivity with other networks like the CON and the dorsal attention network. Additionally, DMN2 and DMN3 are associated with enhanced connections among medial and lateral nodes, respectively.

An alternation between DMN2 and DMN3 would be another instance of the fundamental processing cycle being described throughout the present dissertation, in which prediction error is contested by prediction. The choice of CON5 over CON2 makes sense because it is the only CON NC-state that simultaneously shows a correlation with FPN2, and a correlation-anticorrelation pattern with DMN3 and DMN2, respectively (cf. Figure 2.2). It also shows slightly less correlation with FPN2 than CON2, which may be an NC-state reserved for more effortful cases. We can then imagine the implementation of an input gating strategy as a meta-state temporarily dominated by the MTL subsystem (DMN2), which then transitions into lateral connectivity with DMN3 when searching for the gap, and again settles into DMN2 when finding a match.

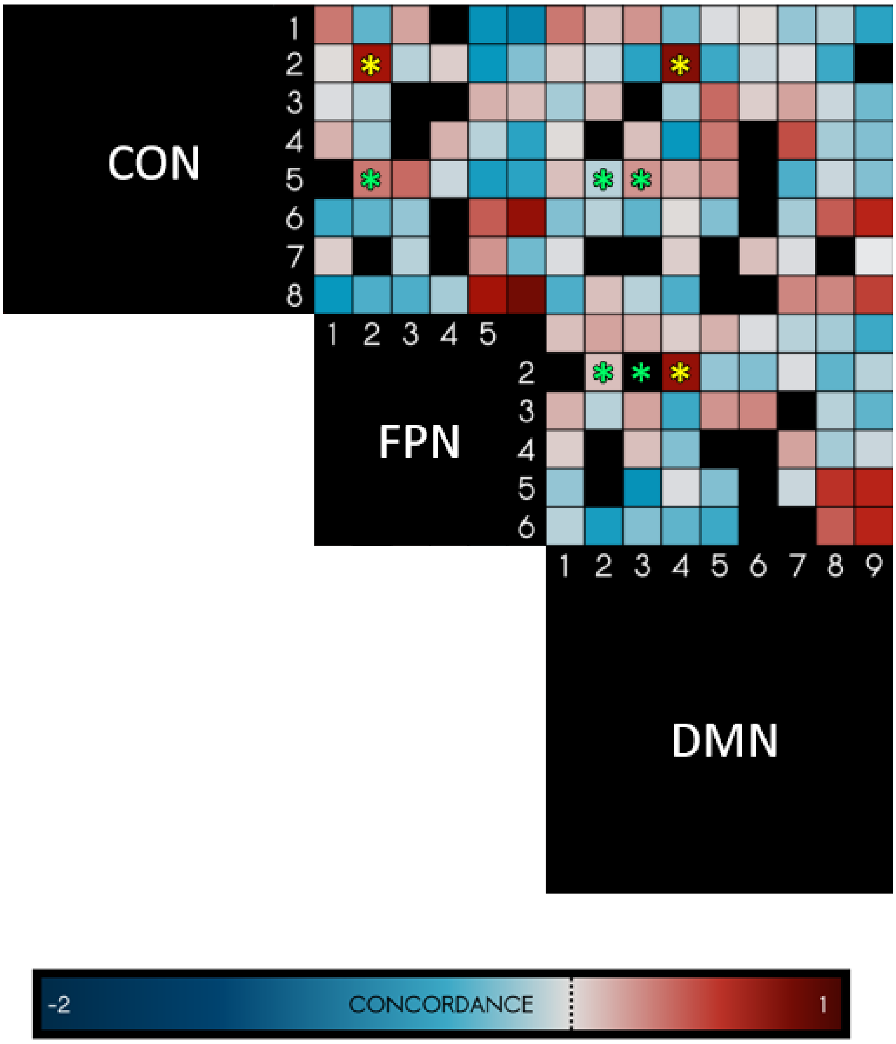


Figure 2.2. Bayesian concordance matrix of the different NC-states of the cingulo-opercular (CON), frontoparietal (FPN), and default-mode networks (DMN). More concordance (in red) entails correlated activity between NC-states, less concordance (in blue) entails more frequent anticorrelations, and the color black represents non-significant concordances (uncorrelated activity, independence). The green asterisks illustrate the proposal of this section that an input gating strategy for long-distance dependencies is implemented by a CON5-FPN2-DMN2/3 meta-state, while a CON2-FPN2-DMN4 meta-state (yellow asterisks) would be reserved for more effortful contexts, perhaps including the output gating strategy. (Adapted from Ciric et al., 2017).

Importantly, this connectivity regime would still allow for local integrations to keep taking place efficiently while the gap is found. The medial core of the DMN is shared by both NC-states, and recent studies show that it is also structured and functions according to a dorsal-ventral partition. In particular, PCC is situated as part of the paralimbic system that links information derived from the internal and external world, serving as an interface between the MTL and higher-level cortex (Mesulam, 1998; Fornito et al., 2012; Braga & Leech, 2015). vPCC and dPCC have been associated with the processing of objects and spatial information, respectively, and seem to act in concert with vmPFC and dmPFC (Vogt et al., 2006; Bzdok et al., 2015). In particular, very effortful states like a working memory maintenance task are associated with the deactivation and modularization of both dPCC and vPCC (Leech & Sharp, 2014; Piccoli et al., 2015), consistent with the old idea that the DMN is task-negative.

However, the CON5-FPN2-DMN3 regime proposed here as a less effortful alternative associated with consolidation may offer a more balanced distribution of activity across the dorsal cingulum (dACC-dPCC), which would allow the efficient integration of intervening materials despite the deployment of a seemingly costly input gating strategy. In support of this, Shenhav et al. (2013) suggest that dACC functions are replicated by more posterior regions like dmPFC: “these regions may serve a similar set of functions applied over a range of abstractness of control signals”, while dACC engagement would be limited to “processes that cannot rely fully on prespecified paramters, that is, that are not automatic”. While intervening words are not automatically processed, they certainly constitute less effortful processing contexts. Finally, it may also be significant that during DMN3 the FPN is uncorrelated with the DMN, as the dorsal stream would have to simultaneously represent the maintenance of the pre-activation and the processing of intervening structures.

In conclusion, it seems that NC-states may potentially delineate the local/complex division from a domain-general, whole-brain perspective, but the difference may still end up being quantitative rather than qualitative, with NC-states that are widely shared across the division, and more contextually determined than what the use of these artificial categories suggests.

2.4. Summary and conclusions

The present chapter has focused on modeling the ventral stream from a predictive coding perspective, contributing to the questions raised by the local/complex syntax debate (Sections 2.1 and 2.1.1). In particular, the function of the ventral stream was described by appealing to multiple patterns of activity that should have a role to play across both members of many of the proposed conceptual divisions in the literature: local and complex syntactic conditions, serial order and order insensitivity, syntax and semantics, production and comprehension, object and event representations, top-down and bottom-up processing, high and low certainty conditions, effortful and automatic processing, left-branching and right-branching grammars, working memory and syntax...

This is only possible because it has been established that the ventral and dorsal streams hold a hierarchical relationship that constitutes a single, domain-general processing system, the FPN-CON-DMN system, in which all linguistic and cognitive functions are implemented by means of dorsal-ventral interactions arising from the constant exchange between feedforward prediction error and feedback predictions.

In Section 2.2.1, this vision was made explicit when the idea that top-down prediction is optional and associated with the dorsal stream was contested on the basis of what we know about the ventral stream being sensitive to prediction, and the experiments that associate increased dorsal stream activity with performance difficulties derived from using a non-optimal precision when solving a task. It was also proposed that predictive coding experiments on the auditory cortex (Blank & Davis, 2016) can be extended to cover the language system as a whole, in order to provide support for this framework.

The second question asked about the specific functions of PFC regions, and the nature and location of integration. Sections 2.2.1 and 2.2.3 proposed that BA 44 engages as a function of allocating precision on lower levels of the rostrocaudal hierarchy, and that its engagement can be associated with both unexpected and expected uncertainty. Only for the latter case its activity is reduced to domain-specific regions and controlled by BA 45, which was defined as the apex of a rostrocaudal hierarchy that departs from posterior regions. The relationship between BA 45 and 44 replicates dorsally the one that exists more generally between the ventral and dorsal streams, and the CON and FPN, and can be summarized like this: the dorsal stream represents the parallel addition of alternative processing routes striving to get into the focus of attention, while the ventral stream stands for the subtraction of these routes by either integrating or discarding among competing alternatives. Sections 2.2.3 and 2.2.6 completed this picture by proposing that the ECFS, like the DMN for the FPN, may act as a necessary counterpart of dorsal activity, with the bulk of the activity displacing towards the ECFS or the FPN more depending on the level of certainty that is present at the moment.

This provides a picture in which integrations are whole-brain agreements similar to the global workspace model of consciousness (Baars, 1997; Dehaene et al., 1998), and is consistent with the fact that independent researchers are finding evidence for the existence of a global workspace in both the FPN (Dehaene, 2014) and the DMN (Vatansever et al., 2015). Ultimately, this hypothesis suggests that the same basic pattern of prediction and prediction error replicates fractally across multiple timescales and levels of analysis, from the interactions between neurons to whole-brain meta-systems that involve the exchange of information across multiple functional networks (Section 2.3).

Section 2.2.2 is another example of this, as it conceives the difference between object and event processing as a continuum that depends on whether whole-brain integrations can be made faster and based on lower-level evidence, or require the employment of longer timescales that allow further integrations and enhanced representational resolutions. This section also shows how the LATL effect (Bemis & Pylkkänen, 2011) may dispel the notion that local syntax is associable with specific patterns of activity and regions, as it proposes that it is specifically the content of local structures (i.e., whether they refer to single entities by using easily accessible words) what arguably drives the observed patterns of activity.

Sections 2.2.4 and 2.2.5 specifically studied how much the proactive capabilities of the ventral stream can be extended without requiring dorsal stream engagement, and what contributions regions like vmPFC and OFC, including BA 47, can make to BA 45. It was argued that the difference between a medial and lateral ventral route may be one based on the complexity or the amount of suggested task sets for their implementation by BA 45, and that vmPFC and specifically the delta oscillation may transmit the inhibition of control mechanisms under conditions of high certainty, just as midline frontal theta transmits the need to exert more cognitive control after an integration.

On the question of whether there is a fundamental difference between local and complex syntax, Sections 2.2.6 and 2.3 have approached the topic from different perspectives, concluding in both cases that it is only a quantitative one. Section 2.2.6 reinterpret the one-to-one association between BA 44 and Merge (Zaccarella & Friederici, 2015) from the point of view of a cognitive system that is actually less engaged with the processing of the stimuli, as it seems to be relying on a lower level of the ventral stream hierarchy. It also provided a hypothesis, compatible with comparative evidence, for why BA 44 engagement is sometimes associated with hierarchy and sometimes is not: this should depend on whether the cognitive system is exploring or exploiting at any particular time.

Section 2.3 showed in turn how what is often identified as complex syntax (an input gating mechanism) is just one domain-general pattern of cross-network interactivity in which a prediction can be maintained while other integrations can proceed independently, and suggested that center-embeddings and the acquisition of long-distance dependencies must actually rely on a different, domain-general meta-system (an output gating mechanism, associated with a reactive ventral stream).

The overall conclusion regarding this topic, which is ultimately why domain-general models are justified (question 4), is that it is possible to show that patterns of activity related to complex syntax can arise during local syntactic conditions, suggesting that the former are a subset of the latter that emerges whenever there is a higher metabolic expenditure, and that the AF is just one small piece of the puzzle of what enables language readiness, constituting an effect of consolidation instead of the cause for Merge or complex syntax. Given that we have managed to show different patterns of activity within both local syntax (ATL-vmPFC vs. BA 47-UF vs. FOP-BA 44) and complex syntax (BA 45-AF/ECFS vs CON-FPN-AF, or CON2-FPN2-DMN4 vs CON5-FPN2-DMN2/3), and that their relationship seems to be a quantitative one in all cases, based on cognitive bandwidth and metabolic expenditure, there does not seem to be any neural basis for the local/complex division.

***

3. Vocal learning and language evolution – a domain-general perspective

This chapter explores the evolutionary relationship between vocal learning and language from a domain-general, bottom-up perspective. It is divided into two parts. The first part critically reviews two hypotheses on the origins on language based on selective pressures acting on the vocal learning component, and on the accidental combination of two distinct vocal domains. It concludes that comparative approaches that start from the domain-specificity of vocal learning cannot tell us much about language evolution. The second part proposes a comparative hypothesis that considers domain-general cognition as the connecting point between vocal learning and language, suggesting that enhancements in the CON-FPN-DMN domain-general architecture necessarily affect the vocal learning component by making it less domain-specific and more connected with the rest of cognition, and ultimately capable of sustaining linguistic functions. In order to make this claim, the evolution of domain-general vocal imitation is discussed as an approximation to the kind of changes that the pre-linguistic ancestors of humans could have experienced, and its neurobiology is compared to the model of language established in the previous two chapters.

3.1. Two domain-specific accounts on the origins of language – why they cannot explain human-specificity

Language is characterized in linguistics as a human-specific trait that consists of the ability to link meaning and sound into hierarchical structures described as mildly context-sensitive (Chomsky, 1963; Joshi, 1985; Hauser et al., 2002). The high degree of specificity in the definition of this trait, the common belief that language is a monolithic module, and the scarcity of archeological, genetic and comparative evidence (Hauser et al., 2014) makes the search for linguistic precursors and evolutionary hypotheses a very difficult and speculative task. This first section critically reviews two evolutionary hypotheses based on avian vocal learning that attempt to contradict this pessimistic outlook, and also states what is still missing from them. The main proposed shortcoming is that comparisons based on selected-for, domain-specific traits fail to account for the domain-general nature of language. The suggested solution is to take a non-modular, bottom-up approach, dividing language into more fundamental, shared components that combine to produce species-specific behaviors after undergoing a process of descent with modification (Poeppel & Embick, 2005; Embick & Poeppel, 2006, 2015; Marcus, 2006; De Waal & Ferrari, 2010). In other words, we would like to be able to explain behavior and compare it across species by studying the workings of the CON-FPN-DMN architecture described before, including the algorithm behind the generation (and development) of specialized systems, as opposed to starting from domain-specific traits like vocal learning and establishing direct bridges with other traits.

This first part is structured as follows. In the next section, the reader is introduced to the concept of vocal learning, including how it compares to human vocal learning and what problems exist when comparing it to language. Section 3.1.2 introduces Okanoya's sexual selection hypothesis, analyzes how it tries to navigate around those problems, and focuses on the shortcomings of selectionist accounts of language origins. Section 3.1.3 introduces Miyagawa et al.'s (2013, 2014) Integration hypothesis and poses challenges for accounts on language origins which, like this one, depart from vocal learning and the phenotype. Section 3.1.4 concludes.

3.1.1. What can vocal learning tell us about language?

Vocal production learning (Petkov & Jarvis, 2012) is the capacity to acquire and modify a repertoire of vocalizations by copying an external model, like a conspecific's song. In birds, it is present in three distantly-related families (songbirds, hummingbirds and parrots), probably indicating a case of convergent evolution (Jarvis, 2004), and it is mainly used to acquire songs for courtship and territorial defense (Hessler & Doupe, 1999). In the bird's brain, vocal learners are distinguished by possessing a clearly defined *song control system* (Nottebohm et al., 1976; Nottebohm, 1991; Jarvis, 2004, 2007, 2017), composed of nuclei (*cores*) that are connected by an anterior pallial-striatal-thalamic loop for the acquisition of songs, and a posterior pallial-syringeal pathway for the production of learned vocalizations. These pathways are situated next to areas dedicated to motor control that activate, for example, when the bird is hopping but not singing, indicating that they may have evolved from them by a process of pathway duplication (Feenders et al., 2008).

The song control system is one of the very few brain structures found that seem to correspond with a specific behavior or cognitive trait on a one-to-one basis: what we normally find is that brain structures are multifunctional and that specific behaviors require distributed activity (Anderson, 2007; Anderson & Finlay, 2013). Such apparent modularity is supported by experiments that selectively impair nuclei and consequently measure qualitative changes in song output (e.g. Scharff & Nottebohm, 1991) that are not reported to co-occur with changes in cognitive performance. This has undoubtedly motivated the study of the song control system as a highly specialized and domain-specific module.

Avian vocal learning is also taken as a precursor to at least some linguistic abilities in humans (what Jarvis, 2007 denominates *spoken language*) given the existence of behavioral, neurocognitive and genetic similarities between vocal learning birds and humans (Jarvis, 2007, 2017), referred next.

The human anterior pathway for speech (or grammar, cf. Ullman, 2001) acquisition is hypothesized to exist within the cortico-striatal-thalamic loops of the brain (layers III and V of neocortex sending collateral projections to striatum and motor cortex), while the human posterior pathway should be traced along layers II and III of the neocortex projecting to layer V in laryngeal motor cortex, which in turn projects to brainstem through nucleus ambiguus (Jarvis, 2004) (Figure 3.1). Studies on genetic expression similarities in these regions support this conclusion, with a special place for the transcription factor FoxP2 and its targets playing a role in both song acquisition and the development of specific-language impairment in humans (Fisher & Scharff, 2009).

Behaviorally, both birds and humans approach the learning of conspecific vocalizations by following a process divided into different stages, with a phase of sensory exposition in which a sound template is memorized, and a *subsong* or babbling period in which vocalizations are matched with that template in a trial-and-error fashion (Brainard & Doupe, 2000). This process, characterized by a functional state of increased cognitive control similar to the ones explored in the previous chapter, is common to all forms of motor learning (Costa, 2011). Also, for both humans and birds the acquisition process is subject to critical or sensitive periods of increased plasticity, and there is idiosyncratic and dialectal variation (Wilbrecht & Nottebohm, 2003).

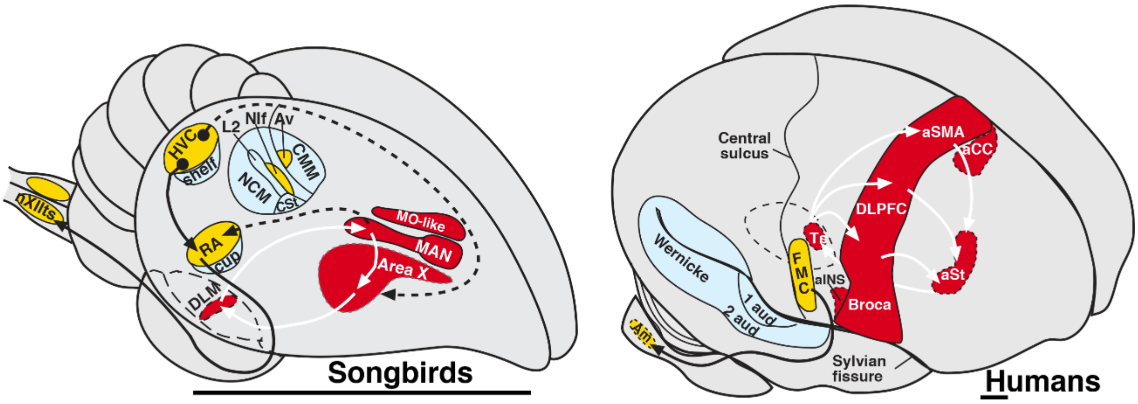


Figure 3.1. Deep homological relationship proposed by Jarvis (2004) to exist between vocal learning birds and humans. Red regions and white arrows represent a prefrontal-striatal-thalamic loop for song/speech acquisition, whereas yellow regions and black arrows represent a direct cortico-laryngeal (pallial-syringeal) pathway for vocal control. Dashed arrows represent connections between both pathways. Blue regions constitute the auditory system. It should be noted that the prefrontal-striatal-thalamic loop runs parallel to regions and pathways involved in domain-general cognition and motor control. Adapted from Jarvis (2004) © Annals New York Academy of Sciences 1016:749-777. PMID: 15313804.

These similarities are used by Jarvis (2004, 2007, 2017) to argue that the neural bases of vocal learning in birds and humans have a deep homological relationship, meaning that both systems may have evolved convergently from a common neural architecture shared by birds and mammals. Note that this hypothesis assumes evolutionary and not just developmental specialization of the domain-general architecture, a process in which natural selection would have played a significant role in both species (Jarvis, 2006).

Beyond that, the neurobiological basis of avian vocal learning is not just similar to this proposed substrate of human vocal learning, but also coincides with mainstream neurocognitive models of language. More specifically, both vocal learning and language are thought to depend on the activity of a domain-specific prefrontal-striatal-thalamic loop (Jarvis, 2004; Jeon et al., 2014), which is mirrored cortically by the activity of the dorsal stream of the frontotemporal network, also responsible for auditory-to-motor transformations (Beecher & Brenowitz, 2005; Hickok & Poeppel, 2007; Bolhuis et al., 2010; Jeon et al., 2014). Given that, and the fact that cognition and sensorimotor processes seem to be more related than what was believed by classic cognitive scientists (e.g., Barsalou, 2010; Pulvermüller & Fadiga, 2010), it is worth asking why attempts to link avian vocal learning and language are not being entirely successful.

The main objections or challenges that comparative hypotheses face when trying to go beyond speech are summarized by Hauser et al. (2014). They argue that song acquisition cannot guide our understanding of language acquisition in humans because "(i) song is a highly specialized and finite system, with the underlying neurobiology linked to one sensory channel (acoustic), and the signal itself is linked to a narrow function and hardly changes once acquired; (ii) when song syllables are combined to create longer structures, there are only limited combinatorial operations and new creations have no impact on the function or "meaning" of the song". In other words, song does not fulfill the functional and structural requirements of language, so it can only inform us about speech or *externalization* (Berwick et al., 2012).

This pessimistic conclusion entails that hypotheses that attempt to link avian vocal learning and language need to face the challenge of explaining the domain-general character of the former, its flexibility, combinatorial character, potentially unbounded uses, and possibility of applying to the vocal component itself, which in most birds seems to be encapsulated from any thought not related to courtship and territorial defense. It also entails that there is no clear way to arrive at the complex structures that characterize language from enhancements in the complexity of vocalizations. More specifically, experiments reveal that animals other than humans have trouble understanding structures beyond finite-state (Chomsky, 1956) (e.g., A^nB^n), whereas humans excel at them, even in non-vocal contexts (Fitch & Hauser, 2004; but cf. also Gentner et al., 2006; Abe & Watanabe, 2011; Beckers et al. 2012).

The hypotheses examined next have attempted to circumvent these objections by appealing to two different strategies: the forces of natural and sexual selection, and emergent behavior resulting from the fortuitous combination of precursor traits. In what follows, we will see why they are still unsatisfactory, and propose an entirely new way of approaching the problem in the second part.

3.1.2. The sexual/relaxed selection hypothesis – shortcomings of selectionist approaches to language evolution

This section analyzes the shortcomings of hypotheses that attempt to explain language evolution from the perspective of selective pressures applying on the vocal component, taking Okanoya (2004, 2012, 2015) as a case study. After introducing the hypothesis, some contrarian arguments will be offered.

Okanoya (2004, 2012, 2015) proposes his sexual/relaxed selection hypothesis for the origins on language drawing from a long tradition dating back to Darwin (1871/2004). The father of natural selection already understood that songbird and language share important features, including a process of vocal learning, and proposed that both fulfilled adaptive roles. He also suggested that sexual selection could have played an important role in how songbird evolved. His argument would be as follows: at some point, females started to gauge a male's health and reproductive potential based on the vocalizations they produced during courtship displays, a costly behavior given that it requires energy and dedication, and exposes the emitter to predators. This would have triggered an arm's race in which, always respecting adaptive constraints, structural complexity would have been favored and gradually increased over generations.

In support of these views, in many bird species the males are the only ones who sing, or male signing abilities outperform those of females, indicating a role for sexual selection as a driving force (Nottebohm & Arnold, 1976). Also, song complexity in males seems to be positively correlated in some species with reproductive success and female preference (Reid et al., 2004). Regardless of the exceptions to these patterns (cf. Catchpole & Slater, 2003), the argument is that learning (and, by extension, a clearly defined song control system) can be the obvious result of sexual selection, which introduced a pressure to make structures more complex until they could no longer be handled by preprogrammed responses and needed to be learned.

In this sense, the song control system can be understood as a peacock's tail: a very complex object that serves a purpose (courtship) that is more ancient than itself, and that produces some unexpected behavioral consequences (the learning and performance of courtship displays). Interestingly, this conception is very similar to how some modular approaches to language (cf. Cook & Newson, 2014 for a review) conceive their object of study: language is also thought to be a very complex object (universal grammar) that serves a purpose that is more ancient than itself (communication or thought), but that produces unexpected behavioral consequences (hierarchical, recursive combination). Given this similarity, could it be that selective pressures driving an increase in the complexity of vocalizations led to the emergence of linguistic structures in humans? Is language just the top of a ladder of computational complexity (Chomsky, 1956), instantiated in a particular brain system (e.g. the frontotemporal network) that experienced a particularly strong performance boost in our ancestors?

Okanoya (2004, 2012, 2015) answers positively to these questions by introducing a new element to this hypothesis: the phenomenon of relaxed selection (Deacon, 2010), which is exemplified by the case of Bengalese finches. These birds are the domesticated variants of white-rumped munias, and present a degree of song complexity and variability that is not matched by their wild counterparts, even in cross-breeding conditions. Their domestication may be the reason why this is the case, as it greatly reduces mate selection, predation pressures, the need to produce a species-specific song for species identification, and high stress levels. These effects should constitute a relaxation of the selective pressures constraining how an adaptive song needs to be, allowing deviations from the norm to flourish. Since Bengalese finch females still have a tendency to select based on song complexity, and relaxed selection gets rid of elements that push in the opposite direction, males tend to increase the complexity of their songs further. Given that there is evidence that humans are a self-domesticated species (Hare et al., 2012; Wilkins et al., 2014), Okanoya argues that language evolution could be a similar case of sexual and relaxed selection combining to produce the kind of complex structures that language requires.

These arguments are problematic from multiple perspectives. First, it should be noted that the relaxed selection hypothesis that Okanoya defends seems to be different from the one originally proposed by Deacon (2010), as the latter did not focus on structural changes only. Deacon (2010) was more concerned with the *demodularization* that the vocal component required in order to connect with domain-general learning. According to him, relaxed selection should open highly pre-specified mechanisms to external (epigenetic) influences, coming both from the individual's brain and the environment: "The result [of relaxed selection] is a tendency to shift control of a previously innate and localized function onto a distributed array of systems that each now only fractionally influence that function. This effectively offloads a significant degree of genetic control onto epigenetic processes, and because of their openness to experiential modification, it increasingly opens the door to the influence of social transmission".

On the other hand, Okanoya's view of relaxed selection has more to do with the conditions that constrain the structures of vocalizations being relaxed, thus allowing selection to target more complex structures. This difference is a crucial one, because there are many ways to increase complexity that do not involve transcending the boundaries of finite-stateness or connecting with meaning whatsoever, for example by introducing variations in acoustic range, repertoire size, or duration, all of them dimensions that could presumably be targeted by females. Moreover, if relaxed selection takes place, and thanks to that the vocal learning component indeed connects with domain-general cognitive capacities, but these remain focused on solving problems related to courtship and territorial defense, an additional explanation is needed for why something like language should ever emerge. Thus, while relaxed selection may be a necessary component of how vocalizations can connect with the rest of cognition more generally, it is clearly not sufficient.

Another problem with applying a unified selectionist perspective to language and vocal learning origins is that this strategy already fails to account for the emergence of vocal learning in birds. For example, Odom et al. (2014) survey the presence of song in the females of many species of birdsongs, reaching the unexpected conclusion that female song is both widespread (present in 71% of the examined species) and ancestral (with the first singing species probably having female singers) (cf. also Langmore, 2000; and Riebel, 2003 for earlier accounts of female song). The authors of this study argue that this may not have been discovered before because avian researchers mostly focus on birds from temperate climates, where harsher climate or predatory conditions may have favored the emergence of gender roles and selection against female song. These findings are interpreted as supporting a more flexible view for the origins of vocal learning than sexual selection, favoring instead social interactions as another important driving force. Notably, the lack of sexual dimorphism is also a problem for sexual selectionist accounts of language evolution, as there does not seem to be strong evidence supporting gender differences in linguistic capacities or traditional language areas in the brain (cf. Wallentin, 2009 for a review).

The existence of a vocal learning continuum that seems to precede selective pressures is also another problem for the hypothesis. The vocal learning continuum hypothesis argues that most complex animals can be considered vocal learners, because this trait does not depend on the presence of a song control system per se, but just on the degree of cortical control an organism can have over its vocalizations (Grimsley et al., 2011; Chabout et al., 2012; Arriaga et al., 2012; Arriaga & Jarvis, 2013; Chabout et al., 2015; cf. Figure 3.2b in Section 3.2.2). For example, despite the belief that mouse vocalizations were innate, mice are still capable of adjusting some acoustic features of their songs depending on the environment they are raised in, possessing dialectal variation (Arriaga et al., 2012).

This framework associates the presence of a song control system with the possibility of acquiring and using more complex vocal structures, and so it finds direct parallelisms with our previous chapters, where the arcuate fasciculus was described as being responsible for the *efficient* (predictive) processing of capabilities that should already exist within a domain-general substrate. This actually tells us that the origin of language may have more to do with the domain-general substrate than with the complexity that vocalizations can actually achieve, whether such complexity is due to selection or relaxed selection. It also tells us that, in the absence of any strong selective pressures targeting vocalizations, there is still a vocal learning component developing at roughly the same pace as other motor domains, its complexity being constrained by the overall metabolic budget and plasticity of a species' brain.

In sum, enhancements in vocal learning abilities can indeed be the product of selection and relaxed selection, but this does not entail that this is the only possible way in which vocal learning abilities can get enhanced, nor that this should necessarily lead to language abilities. Moreover, the claim that avian and human vocal learning are homologous traits does not entail that their evolutionary stories are shared, too. This is easier to notice when we are not talking about cognition: realizing that the genes that control pigmentation are homologous in mammals (Barsh, 1996) does not lead us to suspect that white people, like polar bears, necessarily evolved under the pressures of camouflage.

Clearly, the kind of changes we are looking for to capture the story of vocal and linguistic abilities in humans require more than selection acting on the structure of vocalizations. We will return to these ideas in the second part of this chapter to propose a more promising alternative. Before that, the next section will analyze another hypothesis that also departs from vocal learning to extract conclusions about the origins of language, but does not rely on selection: Miyagawa's Integration hypothesis.

3.1.3. The Integration hypothesis – shortcomings of top-down approaches to language evolution

The Integration hypothesis (Miyagawa et al., 2013, 2014; Miyagawa, 2016) also argues that language evolution departed from vocal learning, but in this case the appeals to selection and relaxed selection are replaced by a non-selectionist explanation. The authors suggest that language may be instead an emergent property of two precursor cognitive systems for communication that got to combine in the human brain for the first time, yielding the possibility of forming linguistic structures. They receive the names of Type E (expressive) and Type L (lexical). An example of type E is birdsong, “which serves to mark mating availability and other ‘expressive’ functions” (Miyagawa et al., 2014), and can create a sustained pattern (a sequence of syllables) that holistically expresses the internal state of the emitter. An example of a type L system is the innately produced alarm calls of some species of primates, which are used to signal the presence and immediate threat of specific predators. Language is said to be unique in possessing properties of both systems: syntactic features like tense and functional words would determine the E-content of phrases, while lexical roots and their referential meanings would be based on L.

The authors speculate that the L system could be inherited from our primate ancestors, whereas E could be either a case of convergent evolution when compared to birdsong, or also coming from primates. Hierarchical structure in language is possible because E and L can alternate, generating phrases in which lexical items are inserted. Regarding how both systems got to combine, the authors speculate that this may be the result of the emergence of shared intentionality in the ancestors of humans (Tomasello et al., 2007), which may have already possessed an L system. They also propose that type E could be initially enclosed in the right hemisphere to compute prosody and rhythm, constituting a musical protolanguage similar to what birds can do. This way of processing would later be extrapolated to the left hemisphere, where dorsal and ventral streams were in charge of the naming abilities of primates (type L), generating a syntactic component.

The Integration hypothesis attempts to do a decomposition of language into more fundamental systems and find precursors for them in other animals. It also tries to anchor its concepts in the brain by considering how they relate to the (purported) roles of the left and right hemispheres, and the capabilities that are attributed to the dorsal and ventral streams.

However, the strategy followed to connect language with evolution is a top-down one (De Waal & Ferrari, 2010), as it departs from language (a highly specific and complex trait) in order to find its precursor for it, as opposed to the bottom-up strategy of looking for more fundamental building blocks of cognition (e.g., prediction-prediction error interactions) that allow us to build such specificity and complexity.

Thus, an initial problem with the Integration hypothesis is that it seems to try to connect one cognitive system to another based on similarities between their phenotypic characteristics. Miyagawa et al. (2014) suggest that language is the combination of expressive and lexical functions, which triggers a search for these functions in other animals, finding that vocal learning and alarm calls resemble them in some aspects. These similarities are judged as good enough, so a model of language evolution is constructed. Unfortunately, if someone is not convinced by the idea that language can be defined in this manner, or that the proposed equivalents are similar enough, a long semantic discussion can ensue in which no attention needs to be paid to the actual mechanisms supporting each trait, and to whether the traits actually cut across natural brain boundaries.

As an example, we can argue that Type L systems do not resemble our lexical capacities because they are not referential at all, as alarm calls are emitted after an immediate perceptual episode, without displacement (if our convictions are different, we can also argue that they are referential because they refer to the internal states of the emitter). We can also look at a recent paper by Price et al. (2015), who analyze the structures of vervet monkey alarm calls (Type L) and conclude that they vary according to the dispositions of the emitter, resembling the holistic, expressive functions of Type E systems, and conclude that vervet monkeys are actually linguistic creatures. In short, comparative hypotheses should not be challenged by behavioral observations, but by new neurobiological findings.

Related to this, there is the additional problem of directly extrapolating cognitive labels to the brain, conflating both levels. For example, Miyagawa et al. (2014) take the song control system of songbirds as the brain equivalent of their Type E system, and speculate that the hierarchical organization of the ventral auditory stream (Rauschecker, 2012), with areas that selectively respond to increasingly more complex auditory objects, may be the neural seat of their Type L system. These direct bridges between cognitive and neurobiological labels are problematic, as they are not supported by the data. We can find animals like corvids, which possess a song control system like other passerines, but lack songs altogether, with their vocalizations being the calls that are characteristic of a Type L system (Clayton & Emery, 2015). If we look at chimpanzees, which lack Type E and L systems entirely, we see that they already possess quantitatively weaker correlates of all the proposed neural equivalents of both systems, including a dorsal/ventral, hierarchical organization of the auditory system, and a direct cortico-laryngeal pathway

(Arriaga & Jarvis, 2013; Bornkessel-Schlesewsky et al., 2015). Also, the hierarchical organization of the ventral stream is not unique to creatures that possess a type L system, but is an integral, conserved property of brain organization (Ahissar & Hochstein, 2004; Kiebel et al., 2008; Wacongne et al., 2011; cf. Chapter 1).

Another problem with top-down approaches is the notion that it is impossible to build cognitive phylogenies (Fitch et al., 2010; Theofanopoulou & Boeckx, 2015). A behavioral label like "sequence learning" can be instantiated by distinct patterns of brain activity and structures across individuals and species (e.g., Heinrich, 1995; Albouy et al., 2013), and the same structures can give rise to very distinct behaviors (e.g. place cells being involved in both navigation and episodic memory, Redish, 1999; domain-general neurons within Broca's region coding for elements of both musical and linguistic processing, Maess et al., 2001; Woolgar et al., 2011). Additionally, the notion of whether a specific behavior exists in an animal often depends on subjective judgments, as they may be present in degrees, and their absence in natural environments does not preclude their emergence in the lab (e.g., Pepperberg, 2006; Zentall et al., 2008), or under different population dynamics (Sapolsky, 2006).

All of this entails that the precursors of specific behaviors do not need to resemble those behaviors at all, and that comparative hypotheses should not be based on cognitive systems but on more elementary neural mechanisms.

Finally, the conceptual units supported by those neural mechanisms may not necessarily match the ones proposed to exist within cognitive systems, but they will probably have a coarser level of granularity (Poeppel & Embick, 2005). This may not be just a limitation of our brain imaging techniques or neuroscientific theories, but entail a problem of incommensurability between different levels of analysis that may require a full reconceptualization of one of the two domains. These two related problems (*granularity mismatch* and *ontological incommensurability*) are proposed by Embick & Poeppel (2006, 2015) to underlie the establishment of any linking hypotheses between cognitive science and neuroscience, and their proposed solution is the development of an interface at the computational level.

In sum, the Integration hypothesis reveals more general problems about top-down approaches to comparative cognition and function to structure mappings. These problems can be avoided by rethinking the hypothesis from a bottom-up, domain-general perspective, in which labels are established after observing recurring patterns of brain activity.

The second part of this chapter uses the insights from the first one in order to establish a new comparative framework for vocal learning and language. Specifically, the elements that will be taken are the role of relaxed selection as a necessary (but not sufficient) requirement for the vocal component to connect with epigenetic influences and domain-general learning, and the need to link vocal learning and language not directly but through the common ground of domain-general neurocognitive mechanisms.

3.2. Domain-general cognition in vocal learning – how vocal imitation could inform the study of language evolution

The previous two hypotheses have in common a conception of vocal learning as a trait that not only precedes language, but acts as the basis for its development. More specifically, both the sexual/relaxed selection and the Integration hypotheses assume that an increase in the structural complexity of vocalizations should be enough to kick-start linguistic abilities, since these are defined on the basis of a high structural complexity (i.e., mildly context-sensitive). This second part of the chapter is focused on providing an alternative hypothesis that is neither constrained by structural requirements only, nor by the idea that complex vocal learning should take us to language. On the contrary, it will be suggested that language can only emerge as part of an enhancement of domain-general cognitive abilities. Such change should apply across domains, and therefore be related to progressive adjustments on the CON-FPN-DMN architecture towards more sophisticated cognitive control abilities, as opposed to being limited to (and triggered by) the vocal modality. Moreover, such change should automatically affect the behavior of all domain-specific systems, including vocal learning, in a *demodularizing* direction, i.e., by greatly expanding the number of behaviors that the organism can carry out by using them.

The remainder of this chapter expands on this hypothesis, and also argues that a comparable evolutionary process has already taken place in the case of *domain-general vocal imitators*, those animal species that are characterized by being both “smart” and capable of vocally imitating species other than their own. If we accept this comparative framework, some new avenues of research appear in the study of language evolution, which are introduced and tentatively explored at the end of this chapter.

The structure of this part is as follows: first, there is a proposal for the specific relationship that exists between vocal learning and linguistic abilities, inspired by how vocal learning and domain-general cognition relate to each other in birds. This includes a more specific comparison with the case of parrots, in which their vocal imitation abilities are argued to result from domain-general cognitive enhancements as opposed to selected-for specializations. Next, these insights are converted into a framework that distinguishes between domain-specific and domain-general cases of vocal learning, which contrasts with the prevalent domain-specific conception that exists in the literature. Next, there is a comparison between the FPN-FTN linguistic relationship studied in the first chapter, and the shell-core distinction in parrots, arguing that they are deep homologies. Finally, the hypothesis that domain-specific vocal learning abilities precede language (the musical protolanguage hypothesis) is explored, concluding that in humans there is enough evidence to favor instead a simultaneous, interdependent evolution of vocal learning and language as part of an enhancement of domain-general cognition.

3.2.1. Vocal imitation as a result of domain-general cognition affecting the vocal learning component

As stated in Section 3.1.1, despite the deep homological relationship between avian vocal learning and the "language network", the current consensus in the literature is that vocal learning is a domain-specific, selected-for trait that is not informative about language origins, but at most about domain-specific concepts like externalization or speech. This position contrasts with the view defended in this dissertation that specializations derive from domain-selective subsets of domain-general systems, inheriting the computational properties of these (Frost et al., 2015; Karmiloff-Smith, 2015), and that human-specific cognition results from quantitative enhancements on a highly conserved architecture. In support of the existence of a deeper connection between language and vocal learning, this section focuses on a variant of vocal learning known as domain-general vocal imitation. This trait is argued to be more transparent when it comes to associating the traits of vocal learning and language under a common substrate, due to an apparently much diminished influence of selective pressures targeting vocalizations. The purpose is not to claim that domain-general vocal imitation is equivalent to language, or even a precursor for it. Instead, it is to show that vocal learning and domain-general cognition are normally highly interdependent traits, and that we should expect that language and speech follow the same pattern.

Domain-general vocal imitation can be defined as the capacity to use the vocal learning component flexibly to learn to replicate sounds from the environment in order to fulfill a large variety of goals. It is observed, for example, in elephants imitating the sounds of a road and human speech (Poole et al., 2005; Stoeger et al., 2012); bottlenose dolphins imitating computer-generated sounds (Richards et al., 1984); beluga whales and other cetaceans imitating human speech (Eaton, 1979; Mercado et al., 2014); fork-tailed drongos imitating predators to scare competitors away from food sources (and even changing to another imitated species so that their victims cannot habituate) (Flower et al., 2014); and, most famously, parrots, with the most clever exemplars being able to learn through a natural method of social transmission the production of English words to refer to meaningful categories like color, shape, material and some degree of numerical competence (Pepperberg, 1981, 1990, 2006, 2010). The fact that some of these behaviors can be spontaneous and not confer any clear evolutionary advantages suggests the possibility that they result from domain-general changes in what an animal is capable of learning (Mercado et al., 2014). In other words, domain-general vocal imitation is not just the capacity to copy other species, but to use the vocal learning component flexibly and for purposes beyond courtship and territorial defense.

This variant of vocal learning should be contrasted with the zebra finch model of vocal learning (Mello, 2014), which dominates the comparative literature with language. As seen in Section 3.1.2, this paradigm states that very strict selective pressures acting on vocalizations dominate all aspects of vocal learning: they determine the existence of a fixed developmental process characterized by copying a specific tutor in a trial-and-error fashion until a song consolidates for sexual maturity and vocal learning stops; they require that only males sing, that songs need to be species-specific and respect dialectal variation, that they are as complex as females prefer, and used in courtship and territorial defense contexts only (Darwin, 1871/2004; Nottebohm & Arnold, 1976; Jarvis, 2006). Any deviation from these constraints faces the risk of maladaptiveness, as the bird can easily be killed by predators or fail to reproduce.

Domain-general vocal imitators defy all of these requirements, showing that vocal learning can exist and develop without them. The most prominent example is parrots. This order contains species that are characterized for being among the most intelligent and socially complex animals, and their learned vocalizations, known as *calls* instead of *songs* (Marler, 2004), are routinely produced by both males and females in all sorts of contexts (Berg et al., 2011; Colbert-White et al., 2014); do not need to be species-specific (Pepperberg, 2010); do not require a trial-and-error process to be produced, but some can be learned from a single exposure (Richards et al., 1984; cf. also Moore, 2004); are open-ended, dependent on a high neuroplasticity existing across domains (Loepelt et al., 2016), and new ones can be learned (with more difficulty) even during old age (Pepperberg, 2010); and do not seem to be easily explained by selectionist hypotheses, with imitations being produced sometimes for no apparent reason at all (Kelley et al., 2008; Kelley & Healy, 2011).

The still untested hypothesis of this section is that vocal learners that possess domain-general vocal imitation capacities should also be characterized by being more intelligent than other related species that do not have them (or that are not capable of them even if pressed in artificial conditions). The reason why this could be the case is that the development of domain-general vocal imitation would require a high metabolic budget without being justified by clear selective advantages, and is therefore more likely to be an epiphenomenon of a brain that can afford to produce this kind of behaviors, i.e., a brain with sufficiently potent domain-general cognitive control mechanisms. These would allow more frequent and extensive periods of exploration as opposed to exploitation, allowing the acquisition of new, more flexible behaviors like vocal imitation. They would also lead to the relaxation of selective pressures across the board, increasing the survival rates of the species.

The objective measurement of intelligence or cognitive control across species that would be ideally needed to test this claim is difficult to conceive, but there are interesting approximations. Lefebvre (2013) reviews a number of studies showing that the rate of innovative behaviors, instances of tool use and social learning tend to go hand in hand across species, hinting at being the effects of domain-general cognition as opposed to independent modular specializations. These tend in turn to be associated with traits like colonization success, longer lifespan, and, more importantly, with measures of relative or absolute brain size. These characteristics are replicated by avian vocal imitators in general, in which a tendency exists for relatively higher abundance and relatively larger brains (Garamszegi et al., 2007). Vocal imitators are also characterized by the traits of short intersong intervals (as opposed to complex vocalizations), and age-dependent expression of repertoires (Garamszegi et al., 2007), meaning that they tend to be open-ended learners that use calls as opposed to songs, and use them often. This may suggest that they possess a more neotenous, plastic brain than domain-specific vocal learners, and that they can afford to vocalize more often and in more contexts without becoming the target of predators (e.g., Hosino & Okanoya, 2000).

Note that these correlations could be meaningful even in the face of exceptions. For example, Benson-Amram et al. (2016) examined correlations between brain size and intelligence by means of setting up the same puzzle boxes task for different animals, and found that the relative brain size of 39 carnivore species from nine families tended to be associated with success in completing this task, while social complexity and manual dexterity did not (cf. MacLean et al., 2014; Stevens, 2014 for similar results about absolute brain size). These results suggest that specific niches can shape domains like manual dexterity or social cognition by individually targeting them and promoting their development, leading to a large variation of cognitive capacities. At the same time, it shows that general brain properties should nonetheless still play an underlying role across the board. Only by testing novel tasks should it be possible to tell these two factors apart.

This is also the reason why the presence or absence of traits should not be the only benchmark that allows the measure of domain-general cognitive capacities, and constitute instead a very rough approximation. Other examples stressing this point are that tool use can appear in insects (Fellers & Fellers, 1976) without implying vocal learning or enhancements in overall intelligence; female singing can appear in some species of birds without implying domain-generality (Odom et al., 2014); and chimpanzees that use learned vocalizations more are known to develop changes in the grey matter distribution of specific regions of their brains (Bianchi et al., 2016), as opposed to global changes that affect other domains of cognition. In short, justifying the connection between domain generality and novel behavior can only be done by examining the neural substrate.

Absolute or relative brain size across species may also be too rough approximations to the kind of change we are looking for to explain an enhancement of domain-general learning capacities (and indeed present exceptions and methodological problems, cf. Healy & Rowe, 2007). Other interesting examples to explore would be functional connectivity patterns (e.g. Cole et al., 2012), brain shape (e.g., Boeckx & Benítez-Burraco, 2014), relative size of frontoparietal cortex (Jung & Haier, 2007; but cf. Barton & Venditti, 2013), cortical thickness (but cf. Yopak et al., 2010), neural packing, and axonal conduction velocities (Dicke & Roth, 2016), etc.

Herculano-Houzel et al. (2014, 2017) also make the case that brain size is far from perfectly correlated with neuronal count and distribution, the latter being more accurate measures of brain function. Related to this, in the case of parrots and corvids, the most intelligent avian species (and also domain-general vocal imitators), we find that their small brains in absolute terms are compensated by highly abundant and densely packed pallial neurons (Dicke & Roth, 2016).

Herculano-Houzel (2012) also introduces the idea that scaling rules affecting neuronal count that exist across the brain of related species (e.g., primates) can give rise to qualitative changes in behavior once a sufficient absolute number of neurons has been reached. One of these scaling rules that seems especially relevant to account for intelligence is instantiated by a progressively greater differentiation of the ratio of neuron density between rostral and caudal areas of the cortex as brain size increases (Finlay & Uchiyama, 2015). Thus, caudal areas have extended periods of neurogenesis in layers 2, 3 and 4 that end up causing a ratio of caudal-to-rostral neuronal density of 1.2 to 1 in rodents, and up to 3 to 1 in the largest brains like that of the baboon. According to the authors, this rule may be especially relevant for frontoparietal cortex, which tends to allometrically scale with the rest of the cortex the most (Kaskan et al., 2005). Simultaneously, the number of spines per layer 3 pyramidal cell increases with brain size towards rostral regions, suggesting that frontal cortex becomes more interconnected with the rest of the brain.

These facts are interpreted by Finlay & Uchiyama (2015) as a correlate of enhancements in cognitive control or domain-general intelligence, as caudal representations become richer and feedforward connections become subject to a more complex hierarchical organization. It should be added that this process also suggests a possible increase in the domain-generality and abstractness of representations, as the output of domain-specific, sensorimotor systems has more chances to converge and interact in the rostrocaudal hierarchy of cognitive control (Badre & D'Esposito, 2009; Nee & D'Esposito, 2016).

It can be proposed that these sort of changes must necessarily affect the vocal component of an organism, both behaviorally and at the brain level. If the vocal learning component was very restricted by the pressures of selection targeting how vocalizations must be to be adaptive, an enhancement of domain-general learning capacities should lead to more frequent exploratory behaviors, increasing the chances that new survival strategies are found and consolidated. This new behavioral variability could act as a form of exogenous redundancy, which Deacon (2010, 2012) suggests is one of the factors producing relaxed selection. Thus, a highly modular vocal learning component would probably experience a demodularization process as a result of domain-general cognitive enhancements, being manifested by its ability to participate in other domains of cognition.

Similarly, if the vocal learning component was not as modularized as in zebra finches, the development of enhanced domain-general cognition should act as a partial shield against modularization, as the evolutionary niche and social organization of the species would become more complex, and preprogrammed responses lose their utility for survival in favor of alternative behaviors that are controlled and flexible.

At the brain level, increases in neuronal count and frontal connectivity should also produce endogenous redundancy, leading to the ontogeny of more functionally complex specialized systems for the consolidation of the newly unlocked exploratory behaviors. In other words, if the bandwidth of the domain-general system increases, cognitive domains should automatically become more functionally sophisticated, as there should be more domain-general material that can differentiate and specialize, and more complex functions that can become subject to specialization through extensive practice[21].

In sum, the argument of this section has been that enhancements in domain-general cognition, supported by multiple potential brain mechanisms, should increase the frequency and complexity of exploratory brain states. This would have the effect of increasing the chances that new behaviors can be found that confer adaptive advantages, increasing the survival rate of the species, and diminishing the importance of very specialized but rigid behaviors. This process amounts to a generation of redundancy (multiple ways of solving the same problem) which would allow selective constraints to relax.

Applied to vocal learning, domain-general enhancements should have similar structural effects to those claimed to occur by Okanoya (2004) in Bengalese finches (cf. Section 3.1.2), but crucially, they would not be limited to them. Instead, they would allow the vocal learning component to participate in the contents of cognition across the board, including newly found behaviors.

The existence of domain-general vocal imitation should be understood in this context, as a behavior that becomes possible once vocalizations can be controlled and

---

[21] Even if we are mainly discussing specializations through the lens of ontogenetic developments arising from extensive practice, this does not discard that complex, controlled, flexibly attained behaviors can turn out to be especially adaptive, leading to selection favoring those individuals that can more efficiently acquire them, and ultimately encoding them as a strong motivational preset (cf. Wheatcroft & Qvarnström, 2015) or a preprogrammed "instinct" (cf. Baldwin, 1896). Note that this chapter is not necessarily against the forces of natural selection shaping some aspects of vocal learning, domain-general cognition and language; the only claim being made is that the zebra finch paradigm assumes the presence of sexual selection targeting vocalizations to increase their structural complexity, and that this is far from being a promising way to explain human vocal learning and linguistic abilities, which may be instead the result of enhancements (selected or not) that affect cognition and motor control across the board.

connected with the rest of cognition without immediately risking maladaptiveness. This explains why no unified selectionist hypothesis seems to be able to account for all instances of vocal imitation observed in nature (Garamszegi et al., 2007), and why vocal imitation does not seem to fit adequately in the more traditional vocal learning paradigms (Mercado et al., 2014).

Finally, this also suggests that the emergence of language and vocal learning could have happened gradually and simultaneously as part of the same set of changes, in which a domain-general system that showed selectivity to different domains improved, producing effects across the board. In this sense, the idea that the vocal component necessarily requires to undergo a process of selection that is different from this set of changes does not seem to be necessary.

In the next section, a new conceptual framework for vocal learning is developed to include all of these insights.

3.2.2. A scale of functional complexity in vocal learning

A current conceptual framework for vocal learning (Petkov & Jarvis, 2012; Arriaga & Jarvis, 2013) posits that convergent evolution departing from a common neural substrate may have led to specialized vocal learning capacities in different species, and claims that a distinction should be made between *vocal production learning* (the ability to learn to produce new vocalizations, mostly a sensorimotor phenomenon), and *auditory sequence learning* (the ability to comprehend auditory sequences of increasing computational complexity). Each ability is not a binary categorical distinction, but is distributed across a continuum of phenotypes (Figure 3.2a-b). The researchers expect that the study of the evolutionary conditions giving rise to different phenotypes can illuminate our understanding of the origins of “spoken language”, a trait that seems to combine the highest known sequence learning skills (the understanding of hierarchical relationships) and one of the highest known vocal production learning skills (due to the presence of a robust direct cortico-laryngeal pathway for motor control, and the possibility of expressing hierarchical relationships by using vocalizations). While it is claimed that auditory sequence learning is a prerequisite for vocal production learning skills, and the possibility of a correlation between both scales is acknowledged, in theory each scale should be independent: we can imagine a vocal non-learner that can understand complex auditory sequences (the authors give the example of a theoretical Turing machine), and an excellent vocal learner that only expresses the most computationally basic structures (in practice, neither of these things tends to happen in nature).

In this section, a third scale will be proposed that captures the domain-general aspects of vocal learning that have been discussed so far, like the existence of domain-general vocal imitation. This scale receives the name of *vocal learning functional complexity scale*, and represents the number of innovative behaviors that can be performed by using the vocal learning component (cf. Figure 3.2c).

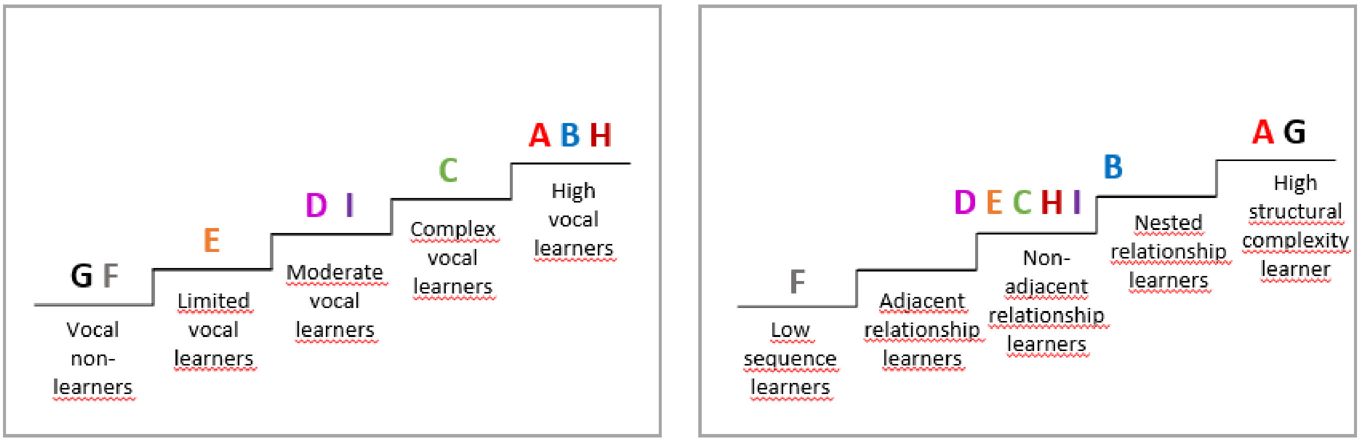
According to the ideas discussed here, this scale should mirror a scale of domain-general learning capacities (intelligence) more accurately than the other two, downplaying the effects of selection pressures targeting domain-specific modalities.

There are two main differences between this scale and the auditory sequence learning scale. While the latter is not committed to the idea that the auditory sequence learning phenotype is based on domain-general cognition, leaving as an open question the reason why an animal may occupy any particular spot within it, the former is. Also, the auditory sequence learning scale proposes a continuum of computational complexity, but in practice uses three discrete categories: low sequence learners (chickens), non-adjacent (finite-state) relationship learners (where virtually all tested animals seem to be, but cf. Wilson et al., 2017), and humans (hierarchy). The functional complexity scale has the advantage of not abiding by structural categories like these, assuming that they can be ultimately explained by domain-general limitations like working memory capacity or cognitive control in some cases, while in others (e.g. songbirds) may have been selected-for. Thus, a more gradual continuum can be proposed by using the functional complexity scale, in which humans would be at the top only as a result of these domain-general dimensions progressively advancing, as opposed to a single dimension of computational complexity advancing independently and abruptly, with advances being due to very diverse causes.

The main difference of the functional complexity scale with respect to the vocal production learning scale is that the latter reflects the degree of cortical control an animal has over its vocalizations (a measure of domain-specific sensorimotor capacities, which can be due either to domain-specific or domain-general enhancements), whereas the former reflects what this control is used for (a function of the degree of enhancement of domain-general cognitive capacities).

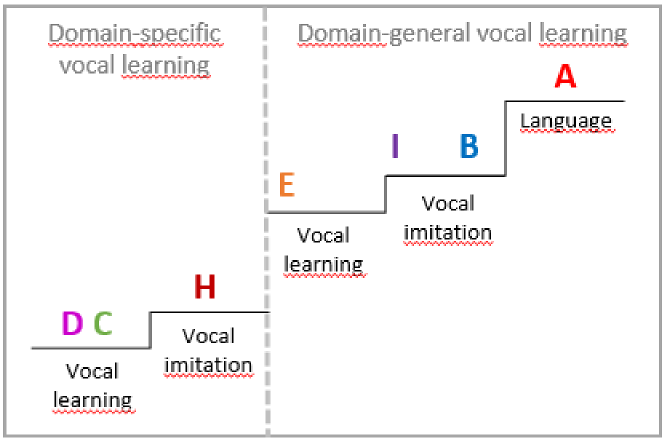
Both dimensions are related to each other in the sense that having enough motor control should be a requirement for increasing the number of potential vocal behaviors. For example, an animal cannot acquire the new function of imitating another species if it cannot modulate its vocal learning component in an imitative direction. However, these two dimensions are far from overlapping: the only thing that is required to occupy a high position in the functional complexity scale is that the vocal learning component can be used flexibly to support new behaviors, regardless of how skillful vocal production capacities are. Accordingly, the reason an imitation is considered “domain-general” is because an *intention* to imitate is recognized, rather than by how accurately the imitation reflects the imitated (Mercado et al., 2014).

In a nutshell, while the first two scales reflect structural enhancements without considering their causes, the scale proposed here organizes phenotypes according to the extent that their vocal learning capacities are interrelated with their domain-general capacities, while arguing that domain-general enhancements push towards increasing such interrelationship.



a) Vocal production learning complexity phenotype **b) Auditory sequence learning phenotype**

[A] Humans	**[B] Parrots**	**[C] Finches**	**[D] Mice**	**[E] Monkeys**
[F] Chickens	**[G] Turing machine**	**[H] Lyrebirds**	**[I] Great apes**	



c) Vocal learning functional complexity phenotype

Figure 3.2. a) A scale that represents the degree of cortical control over vocalizations as a continuum, adapted from Petkov & Jarvis (2012). The horizontal axis represents how structurally or acoustically complex learned vocalizations can be. Parrots (B) and lyrebirds (H) are re-categorized as “high vocal learners” to stress that, regardless of the functional limitations of their vocal learning capacities compared to humans (A), their vocalic sensorimotor skills match and often exceed human capabilities.

b) A scale that represents the computational complexity of the auditory sequences that can be learned in theory by different animals, adapted from Petkov & Jarvis (2012). A theoretical Turing machine is included as an example of a high structural complexity learner with no vocal learning capacities, but the position of animals in this scale seems to roughly correspond with their vocal production learning capacities. Parrots are situated at the very beginning of nested relationship learners because of the Pepperberg studies (cf. Pepperberg, 2010).

c) A scale that represents the number of innovative behaviors that can be supported by vocal learning capacities, which should match a scale of domain-general intelligence or cognitive control. Domain-specific vocal learners may possess heterospecific imitative capacities or not, but they tend to use their vocalizations for courtship and territorial defense mostly, indicating a lower level of intelligence despite the high vocal production complexity of finches and lyrebirds. The dashed line represents a progressive enhancement of domain-general cognitive capacities that has the effect of relaxing general selection pressures, shielding against further domain-specific specialization in the vocal learning component, and greatly increasing its functional complexity. As non-vocal learners, chickens (F) and the theoretical Turing machine (G) are not included. The existence of domain-specific vocal imitation does not indicate that lyrebirds (H) are smarter than mice (D) and finches (C), just that they possess the additional function of vocal imitation. The slightly higher position of "Language" indicates that the change in functional complexity from vocal learning to vocal imitation is a very small one compared to attaining language, but does not necessarily indicate a saltationist scenario.

Here is how the three scales would relate to each other. An increase in auditory sequence learning capacities would correlate with an increase in functional complexity and vocal production learning complexity, provided that it is caused by enhanced domain-general learning capacities. An increase in functional complexity would correlate with increases in auditory sequence learning capacities (unless the effects are too small to jump from one computational phenotype to another, which should happen often) and vocal production learning complexity, since it is caused by enhanced domain-general learning capacities. Finally, an increase in vocal production learning complexity correlates with an increase in auditory sequence learning and functional complexity capacities, but only when it is caused by enhanced domain-general learning capacities. If it is not (e.g., it is the result of sexual selection targeting vocalizations, or extensive practice in the vocal domain), it does not correlate, or auditory sequence learning capacities increase without exceeding finite-state capacities.

Having defined the three scales, we now examine some questions about the functional complexity scale. The first thing to note is that the scale is divided into domain-specific and domain-general vocal learning, with vocal imitation being present in each of them. This classification is due to the existence of *domain-specific vocal imitation*, present in species in which specific selection pressures targeting vocalizations have led to the emergence of imitation (cf., Dalziell et al., 2015 for a selectionist account), or perhaps in which a period of relaxed selection was followed by another period of increased selection. For example, this should be the case of the lyrebird, one of the most outstanding imitators in the animal kingdom, which possesses a song composed almost entirely of sequences of sounds from other species (Kaplan, 1999; Lill & Boesman, 2017). The sequential, trial-and-error character of this behavior, together with its functional similarity to non-imitative birdsong (courtship and territorial defense functions, performed by males), lead us to think that the evolutionary story in this case is not one based on enhanced intelligence, but on the selection pressures acting on vocalizations changing for a species-specific reason that still needs to be determined[22].

---

[22] Another case of domain-specific vocal imitation that seems to deserve its own category is the existence of copying mistakes. Hindmarsh (1986) suggests that, in the case of the European starling, vocal imitation could be a functionless side effect of the vocal learning component accidentally picking up sounds from the acoustic environment of the bird during the process of development. A way of testing whether this is the case for a particular species is to look for whether its mimetic repertoire reflects the most common birds in its acoustic environment (Hamao & Eda-Fujiwara, 2004; Kelley et al, 2008), which only seems to be the case for some species of imitators (cf. Goodale & Kotagama, 2006).

Although the vocal production learning complexity scale by Petkov & Jarvis (2012) has humans in its own category, as “high vocal learners”, this distinction is no longer justified in light of the insights gained from the inclusion of a functional complexity scale. The sensorimotor skills of humans, granted by a robust direct cortico-laryngeal pathway, should not be responsible for their top position in this scale, as we have cases like the lyrebird that seem to greatly surpass human vocal abilities. What makes humans unique is what lies outside of this cortico-laryngeal pathway, such as their vast (potentially unbounded) functional possibilities, belonging to the functional complexity scale.

Another feature of the functional complexity scale is the position that monkeys and great apes occupy within it. The reason why the “E” for primates has been situated on the left-most part of domain-general vocal learning is to acknowledge data from various species of monkeys, for example the phenomenon of learned referential calls in vervet monkeys or tamarins (Struhsaker, 1967; Seyfarth et al., 1980; Kirchhof & Hammerschmidt, 2006), cases in which sequences of sounds can code for single entities and have their components rearranged to change their meaning, which is the method that white-gibbons seem to use for expressing courtship vs warning cospecifics of the presence of predators (Clarke et al., 2006; Fedurek & Slocombe, 2011), or the case of Campbell’s monkeys, which can add a suffix-like modification to an “eagle” call to change its meaning into a more general “arboreal disturbance” call (Ouattara et al., 2009; Fedurek & Slocombe, 2011). We should expect that different species of monkeys are distributed along the continuum of functional complexity that is summarized by the dashed line, with some species fitting into the domain-specific vocal learning category better.

Regarding great apes, these are known to compensate for their limited or moderate vocal production learning capacities in a variety of ways that stress their high intelligence. For example, orangutans modify the frequency of their calls by using their hands and tree leaves to appear bigger and more threatening (Hardus et al., 2008). This complex behavior connects the vocal component with other sensorimotor systems and intentionality, so it justifies the inclusion of these animals as domain-general vocal learners. Additionally, their position as domain-general vocal imitators is supported by a recent experiment that shows that the intention to imitate is present regardless of accuracy: Lameira et al. (2016) demonstrate that orangutans can produce “wookie” calls that instantaneously match human-produced wookies of varying pitch, going beyond their species-typical range and maintaining a turn-taking social setup. Although the initial view was that primates lack vocal plasticity altogether (e.g. Egnor & Hauser, 2004), this view is increasingly being contested especially in the case of great apes, which show vocal production capacities that have been identified as direct precursors of human speech (e.g., Pisanski et al., 2016; Lameira et al., 2016).[23]

In sum, the functional complexity scale reflects the following insight gained in this chapter: that very advanced vocal learning skills do not necessarily entail domain-general learning capacities on par with them. Specializations like the song control system can arise, for example, because sexual selection boosts up domains unevenly by targeting the sequential complexity of songs while leaving the rest of cognition alone. In this case, the

---

[23] In Section 3.2.4, it will be argued that the limited vocal production learning skills of great apes (compared to parrots’ and humans’) may be a natural consequence of scaling rules characteristic of the primate brain. It should be noted that the idea that an enhancement of domain-general intelligence should come associated with enhanced vocal production learning skills does not require that these skills become as sophisticated as those of a parrot or human.

song control system can be considered like a sensorimotor specialization, a localized result of a very domain-specific differentiating force, similar to the barrel fields of rodents (Woolsey & Van der Loos, 1970; Finlay & Uchiyama, 2015). The next section proposes the idea that the song control system of parrots does not fit entirely in this conception, being better explained as the result of a more global differentiating force: an enhancement of domain-general learning capacities affecting the shared neural substrate of vocal learning. This will allow us to claim that parrot vocal learning abilities are more comparable to humans', and speculate on the origins of language in the last section.

3.2.3. Shells as a neurobiological correlate of domain-general vocal learning abilities

This section discusses shells, a recently discovered set of structures that surround the song control system nuclei of parrots, as supporting evidence for the idea that the specialization of the song control system of parrots originates from an enhancement of domain-general cognition, as opposed to sexual selection targeting vocalizations like in the zebra finch case.

Chakraborty et al. (2015) conducted gene expression and connectivity tracing experiments on the song control system of different parrot species, discovering unique features in it. They found that the nuclei within their song control system are divided into cores and surrounding shells, the latter forming anterior and posterior pathways that are adjacent but sparsely interconnected with the cores (which are more ancient and shared with hummingbirds and songbirds). They suggest that the presence of this parallel shell system may be related to parrots' more advanced vocal learning capacities, specifically referring to their vocal imitation abilities. Furthermore, they found that the size of some of the shells, contrarily to what happens for cores, seems to scale allometrically with brain size, and positively correlates with both *vocal and cognitive complexity* in different parrot species.

This discovery seems to support the idea introduced in this chapter that the song control system is affected by changes related to enhanced domain-general learning abilities, in this case an increase of brain size. However, this is not exactly the authors' interpretation: Chakraborty & Jarvis (2015) understand shells as examples of a proposed common principle in brain evolution: the duplication of entire pathways. Just as Feenders et al. (2008) suggested that the song control system duplicated from motor pathways and then acquired new functions in vocal learning, the authors suggest that the shell system duplicated from the original song control system. This duplication would have generated an anatomical redundancy which, due to selective pressures, could then realize new functions without sacrificing the older ones, which still belonged to the core system. Thus, shells are interpreted as new modules which are shaped by selective pressures related to vocal imitation. In other words, they assume that shells allow parrots to ascend in the vocal production learning complexity scale, as opposed to the functional complexity one.

The alternative proposed here is that shells only emerge in cases in which functional complexity increases as a consequence of an enhancement in domain-general cognitive abilities. While the evolutionary mechanism of pathway duplication could be present, the duplicated pathways should not become targets of selection for the fixation of a specific function, but be responsible for quantitative changes in the song control system that facilitate its participation in domain-general behavior. In other words, shells may act as an interface between the song control system and the bird's equivalent of the CON-FPN-DMN architecture. This difference in perspectives should entail the following testable predictions:

First, shells should not be modularly responsible for any new function attributed to parrots' vocal learning abilities, including vocal imitation. We should expect to see, on the contrary, that new functions are a collaboration of the core and shell systems (and, following the previous chapters, with the rest of the brain).

Second, the difference between vocal learning's sensorimotor performance and functionality should be reflected by the shell/core distinction. More specifically, individual differences in the relative volume or degree of activation of cores may be associated with performance differences in the vocal domain (accuracy and frequency of vocalizations, range of different acoustic features, etc.), while individual differences in the relative volume of shells may be associated with measures of intelligence like behavioral innovation. The selective impairment of cores may lead to a more pronounced impairment of the vocal component, which should also affect those behaviors that rely on vocalizations, decreasing both sensorimotor and functional complexity (the latter only in the vocal domain). On the other hand, the selective impairment of shells could have domain-general effects that include vocalizations but are not limited to them, supposing a more generalized cognitive impairment.

This competence/performance distinction would be similar to what happens in language: white matter integrity is a good predictor of language performance in complex syntactic tasks (Skeide et al., 2015), but structural measures are less accurate than brain activity in predicting intelligence (Choi et al., 2008). Also, while domain-general learning capacities like working memory span influence language development (Déak, 2014), there does not seem to be a perfect correlation between domain-general cognitive capacities and degree of consolidation, as measured for example by verbal fluency (Catani et al., 2013; Whiteside et al., 2016).

Related to these two predictions, we already know that the selective impairment of NLc shell (within the parrot's posterior pathway) leads to a decrease in the repertoire size of contact calls (vocalizations that are emitted during diverse social contexts), and an impairment of the equivalent core structure in songbirds leads to a decrease in repertoire size (Pfaff et al., 2007), supporting the collaborative relationship that exists between cores and shells. It would be interesting to compare what happens when cores and shells are impaired in individuals of the same species, and see what consequences it has for domains outside of the acoustic structure of vocalizations.

Third, we should expect to see a gradient of cognitive control (Badre, 2008) from core to shell so that the more cognitively demanding a vocal task is, the more neural activation is recruited in both shells and cores (and other regions outside of the song control system), whereas exploitation is more associated with cores only. This could be tested some day by designing auditory-to-vocal working memory tasks for parrots, provided that a reliable method is found to measure online activity in these structures.

Fourth, it should not be possible to observe activation patterns in cores and shell regions to reliably determine whether they belong to species-specific or heterospecific vocalizations (imitations). The reason is that the fMRI-observable activation patterns of cores and shells depend on how consolidated a vocalization is, not on the source of that vocalization or its acoustic properties. In other words, shells are not modules for heterospecific imitation, they just enable a greater participation of cognitive control processes in the vocal learning component, acting as their periphery (cf. Bassett et al., 2013).

Fifth, shell-like structures will not be present in all animals that can imitate other species, but only in species that have experienced a process of enhancement of domain-general learning capacities, of which an enlargement of brain size or neural count may be approximate neural correlates (but not necessarily the only ones). Alternatively, shells will only be present in species that have experienced two distinct evolutionary stages: one in which vocal learning capacities became extremely adapted as a result of selective pressures without increasing their functional complexity, and one in which the vocal learning component *demodularized* as a result of an enhancement of domain-general learning capacities. In the first case, shells (or some mammalian equivalent, like the frontoparietal network) should be present and active in the most intelligent vocal learning mammals when they use vocalizations for novel functions. In the second case, shells should not be present in mammals, but may be present in corvids, highly intelligent and social songbirds which show domain-general imitative capacities (Emery & Clayton, 2004; Seed et al., 2009; Clayton & Emery, 2015; Plotnik & Clayton, 2015). Crucially, in neither case should shells be present in lyrebirds, despite their outstanding imitative capacities[24].

Sixth, shells, being a necessary component to unlock domain-general cognitive abilities applying to the vocal learning component, should be more anatomically or functionally interconnected with other pallial (cortical) regions than cores. Perhaps hinting at this, Chakraborty et al. (2015) speculate that shells can support dancing or entrainment abilities in parrots, based on the fact that there are limited singing-related specializations of SLIT1, mENK, and TH genes in the PVALB-defined shells, which extend into non-vocal areas of motor regions. This may also be interpreted as evidence that shells possess dynamic coding properties like the FPN (Stokes et al., 2013; cf. Section 1.1).

In sum, these predictions derive from the main idea that the FTN-FPN relationship studied in the first chapter is replicated by the core-shell distinction in parrots. If enough of them were confirmed, they would draw a picture of shells as the brain's response to domain-general cognitive enhancements acting on the shared vocal learning substrate. If this specific evolutionary scenario can be singled out by looking at the brain, it opens up an interesting avenue for doing comparative research with human language that has so far not been explored. This chapter will finish by referring an example of this, analyzing how

---

[24] Some interesting candidates to test these claims would be the humpback whale, which contrarily to other cetaceans does not imitate other species, produces a more complex song, and seems to be less cognitively sophisticated (Mercado et al., 2014); and the marsh warbler, a songbird whose song is almost entirely composed of mimicked bird species (Dowsett-Lemaire, 1979), since there is a closely related species, the sedge warbler, which inhabits a similar environment but displays no imitative capacities at all (Catchpole & Slater, 2003) (according to the present hypothesis, neither of these should have shells).

the previous discussion can help illuminate whether there was a musical protolanguage stage in the human lineage.

3.2.4. Domain-general vocal learning and the musical protolanguage hypothesis

The proposal so far is that the ancestors of both humans and parrots may have experienced a process of enhancement of their domain-general cognitive abilities which necessarily affected their vocal learning components. This view is compatible with two-stage accounts of language origins in which speech came before language, the most prominent being the musical protolanguage hypothesis. However, this section will explain why a one-stage hypothesis is not only as compatible with the previous framework, but also more likely than the musical protolanguage hypothesis. In short, the questions being asked in this section are: can language evolution be accounted for without resorting to a preceding period in which natural selection targeted vocalizations until the cores of a speech control system emerged? And, can we find any evidence for simultaneous evolution in the existence of shells?

The musical protolanguage hypothesis (Darwin, 1871/2004; Fitch, 2006, 2010) claims that the ancestors of humans may have been very capable vocal learners, but not linguistic creatures. In this view, the first stage of language evolution, known as *phonological syntax* (Marler, 1977), would be characterized by the gradual evolution of a selected-for vocal learning component that produces finite-state structures devoid of combinatorial semantics and domain-generality, similar to zebra finch song. Later, a different (still undetermined) change or set of changes would have been responsible for a second stage, leading to the capacity to arrange hierarchical relationships between elements, and to express referential meaning. This stage may have been, according to some researchers, an abrupt one, given the apparently sudden emergence of representational tools in the archeological record coinciding with the emergence of anatomically modern humans, and based on the idea that many aspects of language do not seem to be adaptive (e.g. Chomsky, 1986; Hauser et al., 2002; Tattersall, 2010). Others disagree, conceiving more gradualist perspectives with differing roles for selection (e.g., Pinker & Jackendoff, 2005; Deacon, 2012; Levinson & Holler, 2014).

In any case, the study of vocal learning birds has been interpreted as supporting evidence for the musical protolanguage hypothesis, acting as an example of what the first stage could have been like for humans, based on the known similarities that extend beyond the phenotype. It should be noted, however, that this conclusion is not the only answer that is consistent with the vocal learning data. The alternative defended here is that songbirds took an evolutionary path that the ancestors of humans never took, and that parrots represent a better approximation to what may have happened in our case.

We now know that songbirds and parrots may have shared a common vocal learning ancestor (Suh et al., 2011; Jarvis et al., 2014), and that vocal learning skills are not the result of selection targeting vocalizations but seem to be present in most complex animals (Petkov & Jarvis, 2012). We also know that the ancestor of all songbirds probably had singing females (Odom et al., 2014), meaning that during those times selective pressures targeting vocalizations may have not been as extreme as they currently are in some species like the zebra finch. We can then speculate that the ancestor of songbirds and parrots may have been a limited or moderate vocal learner according to the vocal production complexity scale, and that selective pressures targeting vocalizations made songbirds into more domain-specific vocal learners, increasing their sensorimotor capacities through the development of a shell-less song control system. Meanwhile, parrots underwent a process of increased functional and sensorimotor complexity as a consequence of an enhancement of their domain-general cognition. Therefore, a competing comparative hypothesis is that the ancestors of humans were more like parrots than songbirds, and that a musical protolanguage stage did not exist beyond what all other primates can currently do. In other words, if sensorimotor and functional complexity can evolve simultaneously as a result of the same set of changes, speech and language could have done the same.

Support for this view is found in some considerations about the human posterior pathway, and in the comparison of the FPN-FTN relationship with the shell-core distinction. For the first case, the original view was that humans were unique in having a direct cortico-laryngeal pathway for cortical control of learned vocalizations, similar to the posterior pathway of the song control system in birds (Kuypers, 1958; Zhang et al., 1995; Simonyan & Jürgens, 2003; Jarvis, 2004, 2007, 2017). Non-human primates were later proposed to have a much weaker posterior pathway, consistent with their limited vocal production learning capacities (Arriaga & Jarvis, 2013), which justified their comparison with suboscines, close relatives of songbirds that only possess an incipient posterior pathway (RA nucleus) and lack complex vocal learning abilities (Liu et al., 2013; De Lima et al., 2015).

While the song control system was always considered an offshoot of other domains of motor learning (Feenders et al., 2008), this change of perspective allowed for the first time the possibility that enhancements in sensorimotor complexity taking place outside of the song control cores and the domain of singing could affect the vocal learning component.

A possible way in which this could have happened can be suggested based on a study by Herculano-Houzel et al. (2015). These researchers propose an explanation for the fact that humans, compared to other primates, have a much heavier dependence on motor cortex for all aspects of motor control, including fine-grained movements that in monkeys are performed by spinal cord neurons (Lawrence & Kuypers, 1968). They put forward the idea that brain scaling rules in primates can account for this, as opposed to a human-specific specialization. More specifically, cerebral cortex mass in primates scales linearly with the number of neurons, but spinal cord mass increases disproportionately as it gains neurons. Applying this rule to a larger brain should have the consequence of a dramatic increase in the ratio of motor to spinal cord neurons in favor of the first, which could translate into motor pyramidal neurons having a greater influence in motor control. In a process similar to how the pallial cores of the song control system are also active for unlearned vocalizations (Ter Maat et al., 2014), this corticalization of motor control could entail the emergence of more complex motor plans that can override and take control of the speech control system, increasing its functionality and domain-generality.

The other piece of evidence that supports the idea of a single, continuous stage in humans is the existence in the human brain of a shell-core distinction in the form of the FPN-FTN relationship.

While both the songbird's song control system and the arcuate fasciculus can be described as specializations, they seem to specialize for different purposes. In particular, the FPN may play a larger role in the development and functioning of the latter, enabling the flexibility necessary to acquire and deploy predictions about linguistic input that is much more uncertain and abstract than simple auditory sequences acting as a chunk. If the arcuate fasciculus develops as a result of the consolidation of these processing capacities, this entails that its similarity with the anterior pathway should be smaller than in the case of the posterior pathway with motor cortex, and that its association with domain-general processes should be higher.

Regarding the first prediction, we see that the brain state characterized by dorsal stream engagement is one in which activity seems to be integrating multiple sources, as it is associated with the presence of sound and meaning, and multiple linguistic domains that can hardly be attributed to different pathways or subregions within the stream. These integrative properties of the dorsal stream turn it into a specialization for the efficient processing of domain-general cognitive processes associated with speech, rather than a specialization for sound sequences as in the case of the anterior pathway.

The consequences of this are that the neural similarities between both systems can also be contrasted with some important differences. For example, the lMAN and mMAN cores of the song control system are very localized structures, but their proposed counterparts are a strip extending from anterior insula, anterior cingulate, Broca's region, anterior dorsolateral prefrontal cortex, and anterior presupplementary motor area (Jarvis, 2007), a dorsal stream that not only covers language-specific regions, but others that are among the most functionally diverse of the brain (Anderson et al., 2013), belonging to the CON and FPN. These results can be complemented with a study by Pfenning et al. (2014), in which it is found that the pallial (cortical) regions of the song control systems (lMAN and HVC) are specifically the ones with the least convergent genetic similarities between birds and humans.

Continuing with the underlying genetics, another important difference between the anterior pathway and the dorsal stream that also reinforces the more domain-general character of the latter are recent discoveries on the transcription factor FOXP2[25] (Enard et al., 2002; Zhang et al., 2002; Ullman, 2004; Fisher & Scharff, 2009). Current views (e.g., Berwick & Chomsky, 2011) focus on the association between FOXP2 and impairments in motor control (Vargha-Khadem et al., 1995; Ullman & Pierpont, 2005; Evans et al., 2009) to deny the possibility that it may have anything to do with the faculty of language itself, being relegated instead to its externalization. This view is reinforced by recent findings that some of FoxP2 targets (SLIT/ROBO genes) are involved in the formation of the direct cortico-laryngeal pathway (Wang et al., 2015c).

On the other hand, the human version of FoxP2, when introduced in mice, is known to speed up procedural learning more generally in a maze task (Schreiweis et al., 2014), but does not affect the structure of mouse vocalizations (Hammerschmidt et al., 2015). Hence, a competing interpretation is that FOXP2, by modulating plasticity in cortico-striatal-thalamic loops (Enard et al., 2009), may play a more general role in learning across domains. Based on this, we can speculate that the impact that FOXP2 may have on the specialization of the dorsal stream in humans may be a very indirect one, based on contributing (with many more genes) to initially setting up the extent of the brain’s

---

[25] When fully capitalized, the name of the gene refers to its human-specific variant.

learning capacities. This hints at the possibility that the development of the anterior pathway of the song control system may be more directly genetically controlled than the specialization of the arcuate fasciculus, a hypothesis that would deserve further study.

In sum, the current status of the comparative literature between songbirds and humans does not only show unquestionably relevant points of contact, but also an important difference: a domain-specific/domain-general split that indicates that human vocal learning cores (language-selective regions) are much more complex than the song control system cores of songbirds, even if both rely to an extent in an overlapping neural substrate. The reason for this increased complexity is an enhanced CON-FPN-DMN architecture that is free to explore and consolidate its findings by means of vocal learning, whereas in songbirds the vocal learning process is highly constrained by selection to arrive at simpler, domain-specific solutions.

If the bridges established here between the CON-FPN-DMN architecture and parrot shells, and the corticalization of motor control and the existence of a more robust posterior pathway turn out to be sound, the musical protolanguage hypothesis is no longer needed, and a direct connection between the evolutionary origins of language and speech can be defended.

3.3. Summary and conclusions

The first part of this chapter introduced the idea that attempts to study the connection between vocal learning and language are not promising because they follow top-down, domain-specific, formal or selectionist perspectives on vocal learning. The second part attempted to establish a new framework in which the connection was made from a domain-general perspective, which entailed focusing on how both vocal learning and language are indirectly related through domain-general cognitive capacities. In order to relate vocal learning to domain-general cognitive capacities, it became necessary to disentangle it from the role of sexual selection targeting vocalizations, which seems to produce a very specialized system that does not phenotypically resemble language. The idea of a vocal learning continuum, and the observation that very advanced vocal learning capacities can exist that are less domain-specific, helped in doing that.

Next, it was proposed that an enhancement of domain-general learning capacities should necessarily affect the vocal learning component, either in a demodularizing direction (by means of relaxing selective pressures), or as a shield against new domain-specificity. While it can still be argued that more behavioral variation can lead to more possibilities for specialization, the neural support that was argued to be behind such variation (quantitative enhancements in the CON-FPN-DMN architecture) suggests a larger role for epigenetic influences, flexibility of thought, and more frequent exploration states. Thus, the nature of ontogenetic specializations arising from this state will be a different one than in the case of birdsong, as it would be able to predict more multimodal and complex patterns of activity, like the FTN does in language.

Finally, the case of parrots was studied more closely in order to speculate on language origins. It was argued that the neurobiology of their vocal learning capacities resembles the predictive coding FPN-FTN relationship discussed in the first chapter, and it was suggested that this may not be entirely due to selective pressures acting on vocalizations. On the contrary, it was proposed that the song control system of parrots, and speech and language-specific regions in humans, may have evolved simultaneously with an enhancement of domain-general learning capacities. This view goes against widely accepted views that vocal learning is selected-for independently from the rest of cognition, offers a new perspective on the vocal capacities of parrots (and the phenomenon of vocal imitation as a whole), and introduces an alternative to the musical protolanguage hypothesis by claiming that speech and language can co-evolve.

***

Conclusions and future research

The goal of this last section is not to summarize again the contents of the chapters, but to focus simply on what has been achieved, the hypotheses that have been proposed, and the many gaps that still remain when it comes to providing a coherent domain-general framework of linguistic functions.

Chapter 1 attempted to capture different patterns of activity within the frontotemporal (FTN) and frontoparietal (FPN) networks from the point of view of predictive coding mechanisms, claiming that the system is mediated by a predictive coding hierarchy of dorsal feedforward and ventral feedback connections (Chanes & Barrett, 2016), and a hierarchy of representational complexity that goes from posterior to anterior regions (Badre & D'Esposito, 2009; Nee & D'Esposito, 2016), with the latter being potentially overridden by the former under conditions of higher predictive certainty.

The first half of Chapter 1 started by setting the basis for the predictive processing approach to the brain activation patterns associated with non-adjacent conditions in language, specifically including the relationship between BA 44, the dorsal bank of BA 45, LIFS and SFG. The starting assumptions were that no region fulfilled by itself any particular function, but that they simply reflected the gradual spreading of prediction error as a function of the efficiency of ventral feedback in curbing prediction error.

The main hypothesis can be summarized as follows: LIFS and left-hemispheric domain-general regions within the FPN, including non-language selective subregions of BA 44 and BA 45, should engage as a function of increases in representational complexity under conditions of expected uncertainty. Also, bilateral FPN activity should further engage as unexpected uncertainty increases, leading to overall more distributed activation patterns which may or may not constitute an overall increase of activity.

Thus, the processing of object-first relatives and increases in sentence length are both conditions that allow their proactive parsing, and are therefore associated with the first pattern of activity. Meanwhile, the processing of center-embeddings represents more uncertain conditions, perhaps because these entail the more frequent inhibition among competing alternatives, and would therefore be associated with the second pattern of activity (Fiebach et al., 2004). Thus, the reason why BA 45 is only affected by repetition suppression in the first case (Santi & Grodzinsky, 2010) would be that the second case is an example of processing without any strong priors, i.e., a reactive instead of proactive ventral stream.

The model was put to test by looking at the literature on evoked potentials, which was analyzed in support of a single-stream model of processing in which the dorsal stream is at the service of the ventral stream.

The N400 effect is conceptualized as reflecting the state of the ventral stream at the moment of integrating the bottom-up input, and should vary as a function of how proactively or reactively the stream was operating in the moments before, and also depending on whether the actual bottom-up input confirms or disconfirms its predictions. The P600 effect would reflect the degree of inhibition that is exerted over the dorsal stream, which can be a consequence of how effectively the ventral stream can explain away prediction error at the moment an integration takes place. Finally, the anterior positivity represents a case in which the ventral stream was acting proactively and manages to replace one candidate prediction for another immediately, without requiring any process of evidence accumulation. It is an example of how the predictive coding hierarchy can override the rostrocaudal hierarchy under conditions of higher certainty.

This framework also allowed a domain-general interpretation of the monophasic P600 pattern turning into a biphasic N400+P600 when subjects are given more time to process (Chow et al., 2016). The idea is that events are costlier to integrate than objects, as they require a higher spatiotemporal representational resolution, which is reflected by a necessary increase in the cognitive control hierarchy, and a consequent decrease in certainty (i.e., a more reactive ventral stream). This domain-general effect should translate into language, and be reflected by the monophasic P600. Only the introduction of more time can avoid this tendency, as it would suppose a progressive accumulation of evidence that would allow the pre-activation of the event by the ventral stream, and thus the recovery of the N400+P600 pattern.

If this is not the standard way of operation during perception, it is because its metabolic expenditure is always higher than waiting for the linguistic signal to bias the cognitive system. This is related to the FPN and its association with competing processing alternatives: simulating events (thematic structures) without any strong priors is also present during language production, which is characterized by FPN-FTN synchronization. In order to test the validity of this proposal, an experiment is suggested in which a previous context is provided that allows a more proactive ventral stream: if "the thief managed to get a job at the police station" is followed by "the thief arrested the cop", we may find that the time buffer is no longer needed to detect a biphasic pattern.

Related to the advantage for the processing of single objects over events, another proposal of this framework is that animacy-incongruent sentences like "The student restrained the math problem" (Chow & Phillips, 2013) may shift precision towards lower-level regions, thus being more associated with the P600, while the animacy-congruous "The student hung the math problem" may do the opposite and be more associated with an N400 effect. This speaks about the possibility that very early on during sentence processing the cognitive system can estimate the level of certainty with which it can process, and therefore the patterns of activity that will ensue. It is also consistent with Bever & Poeppel's (2010) analysis-by-synthesis approach to language processing, but not with specific interpretations of it such that either syntax or semantics, or the dorsal or ventral streams should come into play sequentially.

Oscillatory evidence was also looked upon to support this framework: more proactive processing contexts seem to show hippocampal theta increases and cortical alpha decreases prior to stimulus onset (Rommers et al., 2016; Piai et al., 2016). The present framework could contribute to these results by suggesting that alpha decreases reflect increased activity instantiating graded pre-activations. This means that the presence of a pre-activation that is more highly weighted than competing ones may be associated with less overall activity. This would be an alternative hypothesis for why theta power and the positivities associated with disconfirmations become weaker if more activity exists in cortex prior to the stimulus onset (Rommers et al., 2016), and is connected with the idea that the strength of positivities reflects decision certainty (Dröge et al., 2016).

Furthermore, this framework predicts that the N400+P600 pattern, not tested so far, could be associated with larger decreases in pre-stimulus alpha power, and/or with more alpha power at 500-700 ms, reflecting a larger inhibitory effort. Also, it hypothesizes that a monophasic P600 would substitute a larger N400 amplitude if pre-stimulus hippocampal theta and/or neocortical alpha and beta decreases were less significant, and that a role-reversal vs the processing of non-literal meanings (both associated with the monophasic P600) may specifically find variation in this step of alpha/beta power decrease.

Chapter 1 finishes with the neurochemical definitions of expected and unexpected uncertainty, making a series of predictions on the patterns of activity exhibited by the FTN and FPN. Tonic increases of ACh are hypothesized to support FTN engagement, while tonic increases in NE are hypothesized to occur simultaneously during periods of bilateral FPN-FTN coupling. Another possibility is that NE increases phasically during integrations.

Since older populations tend to process long-distance dependencies by relying on more FPN activity, it would be interesting to test whether boosting NE with a drug like idazoxan (Curet et al., 1987) can recover the pattern of increased FTN activity associated with younger populations. The prediction of Section 1.4.5 was that it would because the ERP patterns can be conceived as a continuum, but we can also predict that it would not, because older populations are probably failing to deploy an input gating strategy (cf. Section 2.3) that is possibly dependent on ACh levels only.

Chapter 2 explored brain states characterized by higher certainty, and so it was more focused on the different patterns of ventral stream activity. One of the first conclusions of the chapter stresses a point made in the previous one: that dorsal stream engagement is just one condition in which predictions about upcoming input can manifest themselves, characterized by expected uncertainty and the integration of event representations. Against the view that prediction is optional and dorsal, some examples were provided showing that faster processing latencies can also be associated with ventral stream engagement, and slower ones with dorsal engagement, with the activation patterns ultimately depending on the specifics of the task and the individual.

Rather than suggesting a specific linguistic or predictive function for the dorsal stream (hierarchy, top-down mechanisms), what this indicates is that its activity is better described as reflecting performance changes. A useful approach to capture this was proposed to be the notion of entropy reduction, as it can define the degree of dorsal activity as a function of downward changes in grammatical uncertainty (Nelson et al., 2017b). This would show that the state of the ventral stream (i.e., certainty) needs to be taken into account before attributing functions to the dorsal stream based on its engagement for any particular task, and reinforces the idea that there is a hierarchy.

In relation to the description of the dorsal and ventral streams as predictive, an extrapolation of the results of Blank & Davis (2016) for primary auditory cortex was proposed to apply to the FTN as a whole. The hypothesis was that BA 44 would normally increase its activity as a function of how much a specific input can potentially be processed, or how many different sources of information can potentially be integrated (a proxy for entropy) given the amount of precision allocated to posterior regions, while the ventral stream would engage more as a function of how effectively these integrations can take place (entropy reduction), leading to decreases in dorsal activity. Furthermore, ventral activity would displace towards higher-level regions (ending in BA 45) as prediction error does, provided that expected uncertainty can be maintained.

This perspective helps explain why in Matchin et al. (2017) observe similar BA 44 engagement for both word lists and long sentences, but not for pre-activated sets of two-word phrases: there should be more processing alternatives for the first two cases, while the latter case may have been more effectively integrated, and could be being deployed as a chunk from higher-level regions like BA 47. This would be an example of how the predictive coding hierarchy can override dorsal responses under conditions of high certainty.

The framework can also sketch out the behavior of BA 45: it should engage specifically as a function of higher-level expected uncertainty, attempting to inhibit the activity of posterior dorsal regions like BA 44. Higher-level expected uncertainty here refers to a larger divergence between processing context and long-term memory, which would require a more extensive process of integration, but the question of whether the transition from dorsal BA 44 to dorsal BA 45 is a continuum or reflects the addition of specific mechanisms at the brain level is still pending.

Related to this, in Section 2.2.6 it was argued that the ventral counterpart of BA 44 can be the FOP, a pattern of activity that is more evident when BA 45 is not particularly engaged. This could represent a lower level of the prefrontal cortex hierarchy in action, associated with less representational resolution, less alternative processing routes, less metabolic expenditure or less precision allocated to posterior regions. The balance between FOP and BA 44 engagement would depend on the degree of certainty. The association of the FOP with more localized structures (Friederici, 2012) hints at a division at the cognitive level, but again, mechanistic descriptions are pending, which may as well be simply a continuum.

In sum, BA 44 engagement can be argued to be associated with both expected and unexpected uncertainty, provided that some amount of precision is allocated to posterior regions. Thus, complexity increases in which expected uncertainty can be maintained, like long sentences, should preferentially engage BA 45, but also BA 44, while the opposite pattern could be found in cases of more unexpected uncertainty, simply because BA 45 would be less sensitive to the stimuli.

This framework still requires the specification of the very early mechanisms allocating precision (i.e., estimating certainty) based on the detection of bottom-up input, and an investigation into the thresholds of prediction error that when surpassed would lead to one pattern of activity over another. A suggestion for future work is to look into how this model is replicated by subcortical structures, which should play a role throughout all processing stages, including those happening before the 400 ms mark.

Related to this research goal, it has been suggested here that a domain-general object/event distinction can be an approximation to the kind of task conditions that determine the allocation of precision (Section 2.2.2). This divide was already used to address the monophasic P600 turning into a biphasic pattern above, and here it is investigated more generally, as the ventral predictive hierarchy being able to override dorsal activity under conditions of high certainty. The hypothesis was that the LATL effect, starting in ATL at 200 ms, may be the result of highly certain processing conditions derived from having to process a single entity, which would also cover the processing of a highly consolidated linguistic structure referring to a single entity. This domain-general feature would be also reflected in language developmentally and during processing, with the processing of entities having preference and being associated with a particular activation pattern. The latter would be an example of how local syntactic conditions change their activations depending on content, meaning that a purely structural notion of local syntax is insufficient.

Based on the literature on the LATL effect and the association of the left hemisphere with high certainty and expected uncertainty conditions, an experiment was proposed in which hemispheric cooperation can be blocked in order to see if the LATL effect would still appear, and if it would lead to an intersective instead of collective reading of the sentence “The girls are Finnish and Dutch” (Poortman & Pylkkänen, 2016).

Another one is to what extent tonic lateral conditions depend on activity that is initially medial instead of lateral. This potential collaboration within the processing cycle could be obligatorily occurring in the order of milliseconds, potentially reflecting an analysis-by-synthesis approach, or the ventral suggestion of candidates from long-term memory, and their subsequent adaptation to the processing context by lateral regions (when necessary).

The potential roles of the different ventral pathways were also explored. The conditions that engage the medial route of the UF seem to start from the processing of single entities with the LATL effect but can also cover episodic memory retrievals more generally. The lateral route seems to be more associated with a more complex process: the suggestion of multiple task sets, which would potentially be able to override (inhibit) lateral dorsal activity during conditions of high certainty like the pre-activated sets of two-word phrases studied by Matchin et al. (2017).

At the moment, it is not clear whether a more specific distinction can be proposed, but it has been suggested that the apparent optionality of the UF may have to do with its association with high certainty conditions, which can be replaced by conditions of expected and unexpected uncertainty, associated with slower reaction speeds but more representational resolution (given that BA 45 is the apex of the prefrontal cortex hierarchy)

Another question that requires further investigation is the potential connection between OFC, including BA 47, and the bilateral patterns of activity within the FPN that are uncorrelated with the linguistic signal. This could be how alternative task sets are instantiated when the cognitive system is processing under conditions of expected uncertainty. Contrarily to the increases of activity associated with a multiplication of alternative processing routes, which are known to engage the dorsal stream, these alternative task sets would not actively compete with the current task set whenever the latter is being judged as reliable enough (cf. Donoso et al., 2014). While the role of right hemispheric activity in processing non-literal meanings and more contextual inferences (Burkhardt, 2007; Regel et al., 2011), and in providing less constrained alternatives to pre-activations (Wlotko & Federmeier, 2013) should not be forgotten, this could also be the way in which the brain accomplishes specific serial orders during processing, departing from a set of unordered elements.

A ventrodorsal hierarchy associated with lateral activity (cf. Wilson et al., 2017) was also studied in more detail in Sections 2.2.3 and 2.2.6. It was shown that there is evidence to suggest that the ECFS acts as a necessary ventral counterpart to AF activity across different degrees of certainty, with activity displacing more towards the ECFS as certainty increases and a proactive ventral stream takes control (cf. Rolheiser et al., 2011).

Against the model of the ventral stream by Friederici (2012), it was suggested that the FOP is part of the ECFS instead of the UF. In short, BA 44-FOP would then constitute a lower level of the rostrocaudal hierarchy of cognitive control that ends in BA 45, with FOP occupying a higher position than BA 44 within the predictive coding hierarchy. This allowed us to claim that Zacarella & Friederici (2015) may not have found that Merge is located in the anteroventral part of BA 44, but that phrase structure building can be associated with BA 44/FOP under conditions of low (expected and unexpected) uncertainty (low ACh/NE levels), while the ventral stream is operating reactively and the cognitive system is not allocating a large amount of metabolic resources for processing.

This conclusion stresses why it is necessary to look beyond the dorsal stream to determine what the dorsal stream is actually doing. The activation of BA 44 in monkeys for the processing of adjacent relationships is another argument: if BA 44 is a highly hierarchical sequential processor because it activates for hierarchy in pseudowords, why should it activate at all in monkeys, and do so for the processing of adjacent structures which are more associated with the FOP in humans? The explanation provided here was that the higher cognitive bandwidth of humans should increase overall certainty, leading to more activity displacing towards the ventral region within the BA 44/FOP level of the hierarchy, whereas in monkeys it would be the opposite, leading either to lower expected uncertainty and lower unexpected uncertainty, or to higher unexpected uncertainty than in humans.

An interesting question for future research would be whether the FOP-BA 44 pattern of activity is the one being specifically overridden ventrally by vmPFC under conditions of higher certainty. The potential mechanisms for this have been discussed in Section 2.2.4, with the hypothesis that the delta oscillation following the N400 may be related to ACC or vmPFC inhibiting posterior regions, while the theta oscillation may reflect in turn the need to allocate more precision in posterior, dorsal regions (Cavanagh & Frank, 2014). In light of the inhibitory role of theta, it was also proposed that its increased power at phrase boundaries (Ding et al., 2016) may actually be a reflection of the increased inhibitory needs that the ventral stream would require during less certain conditions, instead of the mechanism that is singlehandedly responsible for phrase structure building.

Chapter 2 finishes with a section that allows us to recast the notion of complex syntax in terms of domain-general patterns of cross-network interactivity, based on recent findings from network neuroscience on the brain basis of cognitive control.

There is the hypothesis that the same syntactic structures can manifest with different patterns of activity depending on factors having to do with prediction, associated with input and output gating strategies (a proactive and reactive ventral stream). There is also the hypothesis that the patterns of activity associated with center-embeddings are not unique to them, but should also exist during the acquisition of long-distance dependencies and their processing as an output gating strategy, again stressing our core idea of a performance-based functional continuum.

These are arguments in favor of considering the AF as localized activations resulting from the possibility of processing under conditions of expected uncertainty contents that are in principle acquired during conditions (and activation patterns) associated with unexpected uncertainty. This localization of activity is the result of a less effortful integration of parallel cognitive domains as a consequence of consolidation (e.g., sound and meaning). Thus, the AF is not the cause of local or complex syntax, but the end developmental result of the language-ready brain.

Additional hypotheses are made regarding the role played by other functional networks during linguistic conditions. It is argued that the dorsal cingulum, encompassing the CON and DMN, probably acts together with the ECFS, constituting an even higher level of the predictive coding hierarchy than it. This is shown by the role of phasic ACC activity in producing network resets, similarly to what has been argued for vmPFC, by tonic ACC activity being associated with the maintenance of task sets, similar to what BA 45 seems to be doing during expected uncertainty, and by the recently discovered collaboration between the DMN and the FPN, which mirrors the collaboration proposed here between the ECFS and AF.

All in all, what this literature suggests is that the whole brain plays a role during language processing. Thus, the different patterns of activity associated with input and output gating strategies were here conceptualized as meta-states, providing the hypothesis that the within and cross-network interactivity patterns of the CON and the DMN should differ across them.

The bottom-line of Chapter 2 is that notions like local and complex syntax represent a continuum of increased activity that is too inadequate to capture all the different regions that can increase their activity both within and across these categories.

Chapter 3 was divided into two parts. The first part critically reviewed two current evolutionary hypotheses on the origins of language through enhancements in vocal learning, and concluded that they were not adequate for multiple reasons. Hypotheses based on sexual and relaxed selection cannot account for the variability of how the vocal learning trait manifests across different species, and the observation that structural enhancements in vocalizations may have taken place does not necessarily lead to the kind of structural enhancements and domain-generality that characterize language. On the other hand, hypotheses based on finding combinations of precursor traits generally suffer from problems related to this top-down approach: determining the presence or absence of a precursor trait in any particular species is to a large extent arbitrary and contextual, since no attention is paid to the neurobiological mechanisms that support them. Besides, when this is done what we find is that traits can be reduced to a common brain language that works across domains, which generates a granularity mismatch problem (Poeppel & Embick, 2005).

The main insights gained from this exercise were that language cannot be defined on the basis of domain-specific structural enhancements, but a connection with domain-general cognition is needed; and that domain-general cognition can in principle evolve at the same pace as motor control and vocal learning abilities, provided that the effects of selection specifically targeting vocalizations are left aside.

The second part proposed to approach the study of vocal learning from the perspective of quantitative enhancements in domain-general cognitive capacities, whether these were selected-for or not. The adopted premise was that domain-general enhancements tend to generate exogenous redundancy (i.e., the possibility of learning more alternative behaviors for survival) and endogenous redundancy (i.e., a frontoparietal network with more neurons available to be retuned or for dynamic coding). These two features are identified by Deacon (2010) as a crucial step in the increased relaxation of the selective pressures acting over an organism, and the increased influence of epigenetic processes, experience, and social transmission over previously highly pre-specified and localized brain modules. Thus, enhancements in domain-general cognition should necessarily affect the vocal learning component in this direction, either by decreasing the highly stringent selective pressures acting over vocalizations, or by preventing them from appearing.

Next, a scale of functional complexity was proposed as a new framework to study the effects of domain-general cognition on vocal learning abilities, which categorizes different species depending on the number of behaviors that can rely on their vocal components, regardless of structural complexity or degree of vocal control. An example of the utility of this framework is that it allows us to understand why vocal imitation tends to correlate with factors like relatively higher abundance, age-dependent expression of repertoires, and larger brains (Garamszegi et al., 2007), but at the same time can also appear in species that are not characterized by their intelligence, like the lyrebird or the marsh warbler.

The main insight gained from this part of Chapter 3 is that there is no need to establish a disconnection between the pace of enhancement of cognitive and motor abilities, including vocal learning abilities, in order to determine the origins of language, unless we have evidence that selective pressures may have played a specific role in separating them. Based on this, a counterargument to the musical protolanguage hypothesis was provided, concluding that the possibility that the evolution of language and speech are intertwined cannot be easily discarded.

Chapter 3 focused as well on the neurobiology behind the vocal learning abilities of parrots, proposing that it reflects their higher position within the functional complexity scale, and that it follows a core-periphery organization that is similar to the frontoparietal and frontotemporal networks in humans. The hypotheses proposed on this topic are all derived from the model established in Chapter 1: shells should be part of a cognitive control hierarchy, not modularly responsible for any function, sensitive to repetition suppression, activated more by effortful exploratory states and uncertainty, and a consequence of enhanced domain-general cognition rather than simply vocal imitation. Also, their selective impairment should have more domain-general consequences than that of cores.

The neurobiology behind the vocal learning abilities of humans was also tackled from this perspective, observing that it reflects a process of corticalization of motor control deriving from primate scaling rules rather than an outcome of new selective pressures, and that the purported equivalent of the human anterior pathway is characterized by its domain-generality, with specializations being a product of consolidation-dependent improvements in processing efficiency, rather than the causes responsible for a speech or language-ready brain.

All in all, this chapter lays the groundwork for establishing a model of animal cognition based on predictive coding mechanisms, in which an intrinsic and highly conserved connection between structure and meaning, like that of sensorimotor systems and perception, is present by design and scales up with cognitive complexity, rather than being an evolutionary novelty characteristic of humans.

A question for future research would be to investigate the ventral aspects of the song control system, with the assumption that domain-specific modules like this, far from being encapsulated, should still form part of meta-states mediated by predictive coding mechanisms. This topic should be more tractable during the subsong stage or in open-ended learners, as we would assume that consolidated sequences displayed during courtship would be characterized by the highest degree of localized activity.

In conclusion, this dissertation has opened new research avenues in multiple domains and perspectives, but its primary goal was simply showing that the mystery of human-specificity can be approached from an interdisciplinary and continuity-based perspective, appealing to factors like predictability, consolidation, reward estimation, individual capacities, and the species-specific cognitive bandwidth.

Human-specificity has been defined as unbounded Merge, or free syntactic combination (Berwick et al., 2013), based on the observation that humans are capable of attaining uniquely complex structural relationships between elements. It has also been explained as being behind the human-specific capacity to combine the outputs of different cognitive modules that are in principle separate, allowing them to simultaneously influence behavioral output (Spelke, 2003). While both perspectives seem to be equivalent at first sight, only the latter is compatible with human cognition and language being the result of quantitative, domain-general changes. The reason for this is that it leaves room for treating the performance effects that content can have on structure building as intrinsic to the operations performed by the cognitive system, as opposed to awkward externalities. In other words, it allows us to argue that the ability to freely combine is not free, but costs a metabolic price, and therefore can be probed empirically and implemented in the brain.

***

References

Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature neuroscience*, *14*(8), 1067-1074.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in cognitive sciences*, *8*(10), 457-464.

Albouy, G., King, B. R., Maquet, P., & Doyon, J. (2013). Hippocampus and striatum: Dynamics and interaction during acquisition and sleep-related motor sequence memory consolidation. *Hippocampus*, *23*(11), 985-1004.

Alday, P. M., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). Towards a computational model of actor-based language comprehension. *Neuroinformatics*, *12*(1), 143-179.

Amso, D., Haas, S., McShane, L., & Badre, D. (2014). Working memory updating and the development of rule-guided behavior. *Cognition*, *133*(1), 201-210.

Amunts, K., & Zilles, K. (2012). Architecture and organizational principles of Broca's region. *Trends in cognitive sciences*, *16*(8), 418-426.

Anderson, M. L. (2007). The massive redeployment hypothesis and the functional topography of the brain. *Philosophical Psychology*, *20*(2), 143-174.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and brain sciences*, *33*(4), 245-266.

Anderson, M. L. (2016). Neural reuse in the organization and development of the brain. *Developmental Medicine & Child Neurology*, *58*(S4), 3-6.

Anderson, M. L., & Finlay, B. L. (2013). Allocating structure to function: the strong links between neuroplasticity and natural selection. *Frontiers in Human Neuroscience*, *7*.

Anderson, M. L., Kinnison, J., & Pessoa, L. (2013). Describing functional diversity of brain regions and brain networks. *Neuroimage*, *73*, 50-58.

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron*, *65*(4), 550-562.

Arriaga, G., & Jarvis, E. D. (2013). Mouse vocal communication system: Are ultrasounds learned or innate? *Brain and Language*, *124*(1), 96-116.

Arriaga, G., Zhou, E. P., & Jarvis, E. D. (2012). Of mice, birds, and men: the mouse ultrasonic song system has some features similar to humans and song-learning birds. *PLoS One*, *7*(10), e46610.

Asato, M. R., Terwilliger, R., Woo, J., & Luna, B. S. (2010). White matter development in adolescence: a DTI study. *Cerebral cortex*, *20*(9), 2122-2131.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, *28*, 403-450.

Auksztulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *cortex*, *80*, 125-140.

Avery, M. C., Nitz, D. A., Chiba, A. A., & Krichmar, J. L. (2012). Simulation of cholinergic and noradrenergic modulation of behavior in uncertain environments. *Frontiers in computational neuroscience*, *6*.

Baars, B. J. (1997). In the theatre of consciousness. Global workspace theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies*, *4*(4), 292-309.

Badre, D. (2008). Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends in cognitive sciences*, *12*(5), 193-200.

Badre, D., & D’Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature reviews. Neuroscience*, *10*(9), 659.

Bahramisharif, A., Jensen, O., Jacobs, J., & Lisman, J. (2017). Serial representation of items during working memory maintenance at letter-selective cortical sites. *bioRxiv*, 171660.

Baldwin, J. M. (1896). A new factor in evolution. *The american naturalist*, *30*(354), 441-451.

Barbas, H. (2015). General cortical and special prefrontal connections: Principles from structure to function. *Annual review of neuroscience*, *38*, 269-289.

Barbas, H., & Rempel-Clower, N. (1997). Cortical structure predicts the pattern of corticocortical connections. *Cerebral cortex (New York, NY: 1991)*, *7*(7), 635-646.

Baron, S. G., & Osherson, D. (2011). Evidence for conceptual combination in the left anterior temporal lobe. *Neuroimage*, *55*(4), 1847-1852.

Baron, S. G., Thompson-Schill, S. L., Weber, M., & Osherson, D. (2010). An early stage of conceptual combination: Superimposition of constituent concepts in left anterolateral temporal lobe. *Cognitive neuroscience*, *1*(1), 44-51.

Barredo, J., Öztekin, I., & Badre, D. (2013). Ventral fronto-temporal pathway supporting cognitive control of episodic memory retrieval. *Cerebral Cortex*, *25*(4), 1004-1019.

Barredo, J., Verstynen, T. D., & Badre, D. (2016). Organization of cortico-cortical pathways supporting memory retrieval across subregions of the left ventrolateral prefrontal cortex. *Journal of neurophysiology*, *116*(3), 920-937.

Barrett, H. C. (2012). A hierarchical model of the evolution of human brain specializations. *Proceedings of the national Academy of Sciences*, *109*(Supplement 1), 10733-10740.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews. Neuroscience*, *16*(7), 419.

Barsalou, L. W. (2010). Grounded cognition: Past, present, and future. *Topics in cognitive science*, *2*(4), 716-724.

Barsh, G. S. (1996). The genetics of pigmentation: from fancy genes to complex traits. *Trends in Genetics*, *12*(8), 299-305.

Barton, R. A., & Venditti, C. (2013). Human frontal lobes are not relatively large. *Proceedings of the National Academy of Sciences, 110*(22), 9001-9006.

Bassett, D. S., Wymbs, N. F., Rombach, M. P., Porter, M. A., Mucha, P. J., & Grafton, S. T. (2013). Task-based core-periphery organization of human brain dynamics. *PLoS computational biology*, *9*(9), e1003171.

Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M., & Grafton, S. T. (2011). Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences*, *108*(18), 7641-7646.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695-711.

Baumgaertner, A., Weiller, C., & Büchel, C. (2002). Event-related fMRI reveals cortical sites involved in contextual sentence integration. *Neuroimage, 16*(3), 736-745.

Beaty, R. E., Benedek, M., Kaufman, S. B., & Silvia, P. J. (2015). Default and executive network coupling supports creative idea production. *Scientific reports*, *5*, 10964.

Beaty, R. E., Benedek, M., Silvia, P. J., & Schacter, D. L. (2016). Creative cognition and brain network dynamics. *Trends in cognitive sciences*, *20*(2), 87-95.

Beckers, G. J., Bolhuis, J. J., Okanoya, K., & Berwick, R. C. (2012). Birdsong neurolinguistics: Songbird context-free grammar claim is premature. *Neuroreport*, *23*(3), 139-145.

Beecher, M. D., & Brenowitz, E. A. (2005). Functional aspects of song learning in songbirds. *Trends in Ecology & Evolution*, *20*(3), 143-149.

Bemis, D. K., & Pylkkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *Journal of Neuroscience*, *31*(8), 2801-2814.

Bemis, D. K., & Pylkkänen, L. (2012). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex*, *23*(8), 1859-1873.

Bemis, D. K., & Pylkkänen, L. (2013). Flexible composition: MEG evidence for the deployment of basic combinatorial linguistic mechanisms in response to task demands. *PloS one*, *8*(9), e73949.

Benson-Amram, S., Dantzer, B., Stricker, G., Swanson, E. M., & Holekamp, K. E. (2016). Brain size predicts problem-solving ability in mammalian carnivores. *Proceedings of the National Academy of Sciences*, *113*(9), 2532-2537.

Berg, K. S., Delgado, S., Cortopassi, K. A., Beissinger, S. R., & Bradbury, J. W. (2011). Vertical transmission of learned signatures in a wild parrot. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20110932.

Berwick, R., & Chomsky, N. (2011). The biolinguistic program: The current state of its evolution and development. *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty*, 19-41.

Berwick, R. C., Beckers, G. J., Okanoya, K., & Bolhuis, J. J. (2012). A bird's eye view of human language evolution. *Frontiers in evolutionary neuroscience*, *4*.

Berwick, R. C., Friederici, A. D., Chomsky, N., & Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends in cognitive sciences*, *17*(2), 89-98.

Betzel, R. F., Fukushima, M., He, Y., Zuo, X. N., & Sporns, O. (2016). Dynamic fluctuations coincide with periods of high and low modularity in resting-state functional brain networks. *NeuroImage*, *127*, 287-297.

Bever, T. G., & Poeppel, D. (2010). Analysis by synthesis: a (re-) emerging program of research for language and vision. *Biolinguistics*, *4*(2-3), 174-200.

Bianchi, S., Reyes, L. D., Hopkins, W. D., Taglialatela, J. P., & Sherwood, C. C. (2016). Neocortical grey matter distribution underlying voluntary, flexible vocalizations in chimpanzees. *Scientific reports*, *6*.

Binder, J. R. (2016). In defense of abstract conceptual representations. *Psychonomic bulletin & review*, *23*(4), 1096-1108.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767-2796.

Binney, R. J., Embleton, K. V., Jefferies, E., Parker, G. J., & Lambon Ralph, M. A. (2010). The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cerebral Cortex*, *20*(11), 2728-2738.

Bland, A. R., & Schaefer, A. (2012). Different varieties of uncertainty in human decision-making. *Frontiers in neuroscience*, *6*.

Blank, H., & Davis, M. H. (2016). Prediction errors but not sharpened signals simulate multivoxel fMRI patterns during speech perception. *PLoS biology*, *14*(11), e1002577.

Blank, I. A., & Fedorenko, E. (2016). Domain-general brain regions do not track linguistic input as closely as language-selective regions. *bioRxiv*, 076240.

Blank, I., Balewski, Z., Mahowald, K., & Fedorenko, E. (2016). Syntactic processing is distributed across the language system. *Neuroimage*, *127*, 307-323.

Blank, I., Kanwisher, N., & Fedorenko, E. (2014). A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *Journal of neurophysiology*, *112*(5), 1105-1118.

Boeckx, C., & Benítez-Burraco, A. (2014). The shape of the human language-ready brain. *Frontiers in psychology*, *5*.

Boeckx, C., Martinez-Alvarez, A., & Leivada, E. (2014). The functional neuroanatomy of serial order in language. *Journal of Neurolinguistics*, *32*, 1-15.

Boisgueheneuc, F. D., Levy, R., Volle, E., Seassau, M., Duffau, H., Kinkingnehun, S., ... & Dubois, B. (2006). Functions of the left superior frontal gyrus in humans: a lesion study. *Brain*, *129*(12), 3315-3328.

Bolhuis, J. J., Okanoya, K., & Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nature reviews. Neuroscience*, *11*(11), 747.

Bonnelle, V., Ham, T. E., Leech, R., Kinnunen, K. M., Mehta, M. A., Greenwood, R. J., & Sharp, D. J. (2012). Salience network integrity predicts default mode network function after traumatic brain injury. *Proceedings of the National Academy of Sciences*, *109*(12), 4690-4695.

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological review*, *113*(4), 787.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Reconciling time, space and function: a new dorsal–ventral stream model of sentence comprehension. *Brain and language*, *125*(1), 60-76.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., ... & Schlesewsky, M. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and language*, *117*(3), 133-152.

Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S. L., & Rauschecker, J. P. (2015a). Neurobiological roots of language in primate audition: common computational properties. *Trends in cognitive sciences*, *19*(3), 142-150.

Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S. L., & Rauschecker, J. P. (2015b). Response to Skeide and Friederici: the myth of the uniquely human 'direct'dorsal pathway. *Trends in cognitive sciences*, *19*(9), 484-485.

Braga, R. M., & Leech, R. (2015). Echoes of the brain: local-scale representation of whole-brain functional networks within transmodal cortex. *The Neuroscientist*, *21*(5), 540-551.

Braga, R. M., Sharp, D. J., Leeson, C., Wise, R. J., & Leech, R. (2013). Echoes of the brain within default mode, association, and heteromodal cortices. *Journal of Neuroscience*, *33*(35), 14031-14039.

Brainard, M. S., & Doupe, A. J. (2000). Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature*, *404*(6779), 762.

Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seiferth, N., Haddad, L., ... & Meyer-Lindenberg, A. (2015). Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proceedings of the National Academy of Sciences*, *112*(37), 11678-11683.

Brennan, J., & Pylkkänen, L. (2012). The time-course and spatial distribution of brain activity associated with sentence processing. *Neuroimage*, *60*(2), 1139-1148.

Brennan, J. R., & Pylkkänen, L. (2017). MEG evidence for incremental sentence composition in the anterior temporal lobe. *Cognitive science*, *41*(S6), 1515-1531.

Brennan, J. R., Stabler, E. P., Van Wagenen, S. E., Luh, W. M., & Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and language*, *157*, 81-94.

Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends in cognitive sciences*, *14*(6), 277-290.

Brilmayer, I., Sassenhagen, J., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2017). Domain-general neural correlates of dependency formation: using complex tones to simulate language. *Cortex*.

Brouwer, H., & Hoeks, J. C. (2013). A time and place for language comprehension: mapping the N400 and the P600 to a minimal cortical network. *Frontiers in Human Neuroscience*, *7*.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. (2017). A Neurocomputational Model of the N400 and the P600 in Language Processing. *Cognitive science*, *41*(S6), 1318-1352.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain research*, *1446*, 127-143.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of cognitive neuroscience*, *5*(1), 34-44.

Bucci, D. J., Holland, P. C., & Gallagher, M. (1998). Removal of cholinergic input to rat posterior parietal cortex disrupts incremental processing of conditioned stimuli. *Journal of Neuroscience*, *18*(19), 8038-8046.

Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. *Nature reviews. Neuroscience*, *13*(5), 336.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and language*, *98*(2), 159-168.

Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *Neuroreport*, *18*(17), 1851-1854.

Bussey, T. J., Wise, S. P., & Murray, E. A. (2002). Interaction of ventral and orbital prefrontal cortex with inferotemporal cortex in conditional visuomotor learning. *Behavioral neuroscience*, *116*(4), 703.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *science*, *304*(5679), 1926-1929.

Bzdok, D., Heeger, A., Langner, R., Laird, A. R., Fox, P. T., Palomero-Gallagher, N., ... & Eickhoff, S. B. (2015). Subspecialization in the human posterior medial cortex. *Neuroimage*, *106*, 55-71.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and brain Sciences*, *22*(1), 77-94.

Catani, M., Mesulam, M. M., Jakobsen, E., Malik, F., Martersteck, A., Wieneke, C., ... & Rogalski, E. (2013). A novel frontal pathway underlies verbal fluency in primary progressive aphasia. *Brain*, *136*(8), 2619-2628.

Catchpole, C. K., & Slater, P. J. (2003). *Bird song: biological themes and variations*. Cambridge: Cambridge University Press.

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in cognitive sciences*, *18*(8), 414-421.

Cavanagh, J. F., Figueroa, C. M., Cohen, M. X., & Frank, M. J. (2011). Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. *Cerebral Cortex*, *22*(11), 2575-2586.

Chabout, J., Sarkar, A., Dunson, D. B., & Jarvis, E. D. (2015). Male mice song syntax depends on social contexts and influences female preferences. *Frontiers in behavioral neuroscience*, *9*.

Chabout, J., Serreau, P., Ey, E., Bellier, L., Aubin, T., Bourgeron, T., & Granon, S. (2012). Adult male mice emit context-specific ultrasonic vocalizations that are modulated by prior isolation or group rearing environment. *PloS one*, *7*(1), e29401.

Chakraborty, M., & Jarvis, E. D. (2015). Brain evolution by brain pathway duplication. *Phil. Trans. R. Soc. B*, *370*(1684), 20150056.

Chakraborty, M., Walløe, S., Nedergaard, S., Fridel, E. E., Dabelsteen, T., Pakkenberg, B., ... & Jarvis, E. D. (2015). Core and shell song systems unique to the parrot brain. *PLoS One*, *10*(6), e0118496.

Chan, D., Fox, N. C., Scahill, R. I., Crum, W. R., Whitwell, J. L., Leschziner, G., ... & Rossor, M. N. (2001). Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Annals of neurology*, *49*(4), 433-442.

Chanes, L., & Barrett, L. F. (2016). Redefining the role of limbic areas in cortical processing. *Trends in cognitive sciences*, *20*(2), 96-106.

Chang, C., & Glover, G. H. (2010). Time–frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage, 50*(1), 81-98.

Chatham, C. H., & Badre, D. (2015). Multiple gates on working memory. *Current opinion in behavioral sciences, 1*, 23-31.

Chatham, C. H., Frank, M. J., & Badre, D. (2014). Corticostriatal output gating during selection from working memory. *Neuron, 81*(4), 930-942.

Chen, T., Cai, W., Ryali, S., Supekar, K., & Menon, V. (2016). Distinct global brain dynamics and spatiotemporal organization of the salience network. *PLoS Biol, 14*(6), e1002469.

Choi, Y. Y., Shamosh, N. A., Cho, S. H., DeYoung, C. G., Lee, M. J., Lee, J. M., ... & Lee, K. H. (2008). Multiple bases of human intelligence revealed by cortical thickness and neural activation. *Journal of Neuroscience, 28*(41), 10323-10329.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on information theory, 2*(3), 113-124.

Chomsky, N. (1963). Formal properties of grammars. In R. D. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2). New York: Wiley, 1963.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.

Chomsky, N. (1995/2014). *The minimalist program*. MIT press.

Chow, W. Y. 2013. *The temporal dimension of linguistic prediction*. College Park, MD: University of Maryland, College Park dissertation.

Chow, W. Y., & Phillips, C. (2013). No semantic illusions in the “Semantic P600” phenomenon: ERP evidence from Mandarin Chinese. *Brain research, 1506,* 76-93.

Chow, W. Y., Momma, S., Smith, C., Lau, E., & Phillips, C. (2016). Prediction as memory retrieval: timing and mechanisms. *Language, Cognition and Neuroscience, 31*(5), 617-627.

Chow, W. Y., Smith, C., Lau, E., & Phillips, C. (2015). A “bag-of-arguments” mechanism for initial verb predictions. *Language, Cognition and Neuroscience, 31*(5), 577-596.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences, 39.*

Christiansen, M. H., & Müller, R. A. (2015). Cultural recycling of neural substrates during language evolution and development. *The cognitive neurosciences V*, 675-682.

Chrysikou, E. G., Weber, M. J., & Thompson-Schill, S. L. (2014). A matched filter hypothesis for cognitive control. *Neuropsychologia*, *62*, 341-355.

Ciric, R., Nomi, J. S., Uddin, L. Q., & Satpute, A. B. (2017). Contextual connectivity: A framework for understanding the intrinsic dynamic architecture of large-scale functional brain networks. *Scientific Reports*, *7*.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181-204.

Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *PloS one*, *1*(1), e73.

Clayton, N. S., & Emery, N. J. (2015). Avian models for human cognitive neuroscience: a proposal. *Neuron*, *86*(6), 1330-1342.

Cloutman, L. L. (2013). Interaction between dorsal and ventral processing streams: where, when and how? *Brain and Language*, *127*(2), 251-263.

Cocchi, L., Sale, M. V., Lord, A., Zalesky, A., Breakspear, M., & Mattingley, J. B. (2015). Dissociable effects of local inhibitory and excitatory theta-burst stimulation on large-scale brain dynamics. *Journal of neurophysiology*, *113*(9), 3375-3385.

Cocchi, L., Zalesky, A., Fornito, A., & Mattingley, J. B. (2013). Dynamic cooperation and competition between brain systems during cognitive control. *Trends in cognitive sciences*, *17*(10), 493-501.

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1481), 933-942.

Colbert-White, E. N., Corballis, M. C., & Fragaszy, D. M. (2014). Where apes and songbirds are left behind: A comparative assessment of the requisites for speech. *Comparative Cognition & Behavior Reviews*, *9*.

Cole, M. W., & Schneider, W. (2007). The cognitive control network: integrated cortical regions with dissociable functions. *Neuroimage*, *37*(1), 343-360.

Cole, M. W., Yarkoni, T., Repovš, G., Anticevic, A., & Braver, T. S. (2012). Global connectivity of prefrontal cortex predicts cognitive control and intelligence. *Journal of Neuroscience*, *32*(26), 8988-8999.

Collins, A., & Koechlin, E. (2012). Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biol*, *10*(3), e1001293.

Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic bulletin & review*, *8*(2), 331-335.

Cook, V., & Newson, M. (2014). *Chomsky's universal grammar*. John Wiley & Sons.

Cools, R. (2015). The cost of dopamine for dynamic cognitive control. *Current Opinion in Behavioral Sciences*, *4*, 152-159.

Cools, R., & D'Esposito, M. (2011). Inverted-U–shaped dopamine actions on human working memory and cognitive control. *Biological psychiatry*, *69*(12), e113-e125.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews. Neuroscience*, *3*(3), 201.

Costa, R. M. (2011). A selectionist account of de novo action learning. *Current opinion in neurobiology*, *21*(4), 579-586.

Coulson, S. (1998). ERPs and domain specificity: Beating a straw horse. *Language and cognitive processes*, *13*(6), 653-672.

Curet, O., Dennis, T., & Scatton, B. (1987). Evidence for the involvement of presynaptic alpha-2 adrenoceptors in the regulation of norepinephrine metabolism in the rat brain. *Journal of Pharmacology and Experimental Therapeutics*, *240*(1), 327-336.

Cutini, S., Scatturin, P., Menon, E., Bisiacchi, P. S., Gamberini, L., Zorzi, M., & Dell'Acqua, R. (2008). Selective activation of the superior frontal gyrus in task-switching: an event-related fNIRS study. *NeuroImage*, *42*(2), 945-955.

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual review of psychology*, *66*.

Dahmani, L., & Bohbot, V. D. (2015). Dissociable contributions of the prefrontal cortex to hippocampus-and caudate nucleus-dependent virtual navigation strategies. *Neurobiology of learning and memory*, *117*, 42-50.

Dalziell, A. H., Welbergen, J. A., Igic, B., & Magrath, R. D. (2015). Avian vocal mimicry: a unified conceptual framework. *Biological Reviews*, *90*(2), 643-668.

Danielmeier, C., Eichele, T., Forstmann, B. U., Tittgemeyer, M., & Ullsperger, M. (2011). Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas. *The Journal of Neuroscience*, *31*(5), 1780-1789.

Darwin, C. (1871/2004). *The Descent of Man*. London: Penguin.

De Lima, J. L., Soares, F. A., Remedios, A., Thom, G., Wirthlin, M., Aleixo, A., ... & Schneider, P. N. (2015). A putative RA-like region in the brain of the scale-backed antbird, Willisornis poecilinotus (Furnariides, Suboscines, Passeriformes, Thamnophilidae). *Genetics and molecular biology*, *38*(3), 249-254.

De Waal, F. B., & Ferrari, P. F. (2010). Towards a bottom-up perspective on animal and human cognition. *Trends in cognitive sciences*, *14*(5), 201-207.

Deacon, T. W. (2010). A role for relaxed selection in the evolution of the language capacity. *Proceedings of the National Academy of Sciences*, *107*(Supplement 2), 9000-9006.

Deacon, T. W. (2012). Beyond the symbolic species. In *The symbolic species evolved* (pp. 9-38). Springer Netherlands.

Deák, G. O. (2014). Interrelations of language and cognitive development. In P. Brooks & V. Kampe (Eds.). *Encyclopedia of Language Development.* SAGE.

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin.

Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70*(2), 200-227.

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, *56*(2), 384-398.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, *79*(1), 1-37.

Dehaene, S., Kerszberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences*, *95*(24), 14529-14534.

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron*, *88*(1), 2-19.

Dehaene-Lambertz, G., & Spelke, E. S. (2015). The infancy of the human brain. *Neuron*, *88*(1), 93-109.

Del Prato, P., & Pylkkänen, L. (2014). MEG evidence for conceptual combination but not numeral quantification in the left anterior temporal lobe during language production. *Frontiers in psychology, 5*.

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia, 61*, 150-162.

DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences, 109*(8), E505-E514.

DeWitt, I., & Rauschecker, J. P. (2013). Wernicke's area revisited: parallel streams and word processing. *Brain and language, 127*(2), 181-191.

Dick, A. S., Bernal, B., & Tremblay, P. (2014). The language connectome: new pathways, new concepts. *The Neuroscientist, 20*(5), 453-467.

Dicke, U., & Roth, G. (2016). Neuronal factors determining high intelligence. *Phil. Trans. R. Soc. B, 371*(1685), 20150180.

Dien, J., Michelson, C. A., & Franklin, M. S. (2010). Separating the visual sentence N400 effect from the P400 sequential expectancy effect: Cognitive and neuroanatomical implications. *Brain research, 1355*, 126-140.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature neuroscience, 19*(1), 158.

Dixon, M. L., Andrews-Hanna, J. R., Spreng, R. N., Irving, Z. C., Mills, C., Girn, M., & Christoff, K. (2016). Interactions between the default network and dorsal attention network vary across default subsystems, time, and cognitive states. *NeuroImage*.

Domenech, P., & Koechlin, E. (2015). Executive control and decision-making in the prefrontal cortex. *Current opinion in behavioral sciences, 1*, 101-106.

Donchin, E. (1981). Surprise!… surprise? *Psychophysiology, 18*(5), 493-513.

Donoso, M., Collins, A. G., & Koechlin, E. (2014). Foundations of human reasoning in the prefrontal cortex. *Science, 344*(6191), 1481-1486.

Dosenbach, N. U., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., ... & Petersen, S. E. (2006). A core system for the implementation of task sets. *Neuron, 50*(5), 799-812.

Dosenbach, N. U., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in cognitive sciences, 12*(3), 99-105.

Dowsett-Lemaire, F. (1979). The imitative range of the song of the marsh warbler Acrocephalus palustris, with special reference to imitations of African birds. *Ibis, 121*(4), 453-468.

Dröge, A., Fleischer, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2016). Neural mechanisms of sentence comprehension based on predictive processes and decision certainty: electrophysiological evidence from non-canonical linearizations in a flexible word order language. *Brain research, 1633,* 149-166.

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition, 92*(1), 145-177.

Dubis, J. W., Siegel, J. S., Neta, M., Visscher, K. M., & Petersen, S. E. (2016). Tasks driven by perceptual information do not recruit sustained BOLD activity in cingulo-opercular regions. *Cerebral Cortex, 26*(1), 192-201.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in cognitive sciences*, *14*(4), 172-179.

Duncan, J. (2013). The structure of cognition: attentional episodes in mind and brain. *Neuron*, *80*(1), 35-50.

Durstewitz, D., & Seamans, J. K. (2008). The dual-state theory of prefrontal cortex dopamine function with relevance to catechol-o-methyltransferase genotypes and schizophrenia. *Biological psychiatry*, *64*(9), 739-749.

Eaton, R. L. (1979). A beluga whale imitates human speech. *Carnivore*, *2*, 22-23.

Egnor, S. R., & Hauser, M. D. (2004). A paradox in the evolution of primate vocal learning. *Trends in neurosciences*, *27*(11), 649-654.

Embick, D., & Poeppel, D. (2006). Mapping syntax using imaging: problems and prospects for the study of neurolinguistic computation. *Encyclopedia of language and linguistics*, *2*, 484-486.

Embick, D., & Poeppel, D. (2015). Towards a computational (ist) neurobiology of language: correlational, integrated and explanatory neurolinguistics. *Language, cognition and neuroscience*, *30*(4), 357-366.

Emery, N. J., & Clayton, N. S. (2004). The mentality of crows: convergent evolution of intelligence in corvids and apes. *science*, *306*(5703), 1903-1907.

Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S. M., Blass, T., Somel, M., ... & Becker, L. (2009). A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell*, *137*(5), 961-971.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S., Wiebe, V., Kitano, T., ... & Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, *418*(6900), 869-872.

Eriksson, J., Vogel, E. K., Lansner, A., Bergström, F., & Nyberg, L. (2015). Neurocognitive architecture of working memory. *Neuron*, *88*(1), 33-46.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2), 321-335.

Fair, D. A., Cohen, A. L., Dosenbach, N. U., Church, J. A., Miezin, F. M., Barch, D. M., ... & Schlaggar, B. L. (2008). The maturing architecture of the brain's default network. *Proceedings of the National Academy of Sciences*, *105*(10), 4028-4032.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491-505.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of memory and Language*, *41*(4), 469-495.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and language*, *115*(3), 149-161.

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in psychology*, *5*.

Fedorenko, E., & Thompson-Schill, S. L. (2014). Reworking the language network. *Trends in cognitive sciences*, *18*(3), 120-126.

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, *108*(39), 16428-16433.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca’s area. *Current Biology*, *22*(21), 2059-2062.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, *110*(41), 16616-16621.

Fedorenko, E., Hsieh, P. J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *Journal of neurophysiology*, *104*(2), 1177-1194.

Fedurek, P., & Slocombe, K. E. (2011). Primate vocal communication: a useful tool for understanding human speech and language evolution? *Human Biology*, *83*(2), 153-173.

Feenders, G., Liedvogel, M., Rivas, M., Zapka, M., Horita, H., Hara, E., ... & Jarvis, E. D. (2008). Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin. *PLoS One*, *3*(3), e1768.

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in human neuroscience, 4.*

Fellers, J. H., & Fellers, G. M. (1976). Tool use in a social insect and its implications for competitive interactions. *Science*, *192*(4234), 70-72.

Fengler, A., Meyer, L., & Friederici, A. D. (2016). How the brain attunes to sentence processing: Relating behavior, structure, and function. *NeuroImage*, *129*, 268-278.

Fernández, T., Harmony, T., Mendoza, O., López-Alanís, P., Marroquín, J. L., Otero, G., & Ricardo-Garcell, J. (2012). Event-related EEG oscillations to semantically unrelated words in normal and learning disabled children. *Brain and cognition*, *80*(1), 74-82.

Ferreira, F., & Patson, N. D. (2007). The 'good enough'approach to language comprehension. *Language and Linguistics Compass*, *1*(1-2), 71-83.

Ferstl, E. C., Neumann, J., Bogler, C., & Von Cramon, D. Y. (2008). The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Human brain mapping*, *29*(5), 581-593.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of psycholinguistic research*, *30*(3), 321-338.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, *47*(2), 250-272.

Fiebach, C. J., Schlesewsky, M., Lohmann, G., Von Cramon, D. Y., & Friederici, A. D. (2005). Revisiting the role of Broca's area in sentence processing: syntactic integration versus syntactic working memory. *Human brain mapping*, *24*(2), 79-91.

Finlay, B. L., & Uchiyama, R. (2015). Developmental mechanisms channeling cortical evolution. *Trends in neurosciences*, *38*(2), 69-76.

Fisher, S. E., & Scharff, C. (2009). FOXP2 as a molecular window into speech and language. *Trends in Genetics*, *25*(4), 166-177.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, *100*(1), 173-215.

Fitch, W. T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*(5656), 377-380.

Fitch, W., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, *1316*(1), 87-104.

Fitch, W. T., Huber, L., & Bugnyar, T. (2010). Social cognition and the evolution of language: constructing cognitive phylogenies. *Neuron*, *65*(6), 795-814.

FitzGerald, T. H., Dolan, R. J., & Friston, K. J. (2014). Model averaging, optimal inference, and habit formation. *Frontiers in human neuroscience, 8*.

Flower, T. P., Gribble, M., & Ridley, A. R. (2014). Deception by flexible alarm mimicry in an African bird. *Science*, *344*(6183), 513-516.

Fornito, A., Harrison, B. J., Zalesky, A., & Simons, J. S. (2012). Competitive and cooperative dynamics of large-scale brain functional networks supporting recollection. *Proceedings of the National Academy of Sciences*, *109*(31), 12788-12793.

Frank, M. J., & Badre, D. (2011). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: computational analysis. *Cerebral cortex*, *22*(3), 509-526.

Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends in cognitive sciences*, *13*(4), 175-181.

Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in cognitive sciences*, *16*(5), 262-268.

Friederici, A. D. (2017). Evolution of the neural language network. *Psychonomic bulletin & review*, *24*(1), 41-47.

Friederici, A. D., & Gierhan, S. M. (2013). The language network. *Current Opinion in Neurobiology*, *23*(2), 250-254.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(7), 2458-2463.

Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Human brain mapping*, *31*(3), 448-457.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive brain research*, *1*(3), 183-192.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*(1456), 815-836.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127-138.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1521), 1211-1221.

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biological cybernetics*, *102*(3), 227-260.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in cognitive sciences*, *19*(3), 117-125.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *Center for the Study of Reading Technical Report; no. 257*.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, *440*(7088), 1204.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1-76.

Gallagher, A., Béland, R., Vannasing, P., Bringas, M. L., Sosa, P. V., Trujillo-Barreto, N. J., ... & Lassonde, M. (2014). Dissociation of the N400 component between linguistic and non-linguistic processing: A source analysis study. *World Journal of Neuroscience*, *4*(01), 25.

Gambi, C., Pickering, M. J., & Rabagliati, H. (2016). Beyond associations: Sensitivity to structure in pre-schoolers' linguistic predictions. *Cognition*, *157*, 340-351.

Garamszegi, L. Z., Eens, M., Pavlova, D. Z., Avilés, J. M., & Møller, A. P. (2007). A comparative study of the function of heterospecific vocal mimicry in European passerines. *Behavioral Ecology*, *18*(6), 1001-1009.

Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of memory and language*, *58*(2), 161-187.

Gerlach, K. D., Spreng, R. N., Gilmore, A. W., & Schacter, D. L. (2011). Solving future problems: default network and executive activity associated with goal-directed mental simulations. *Neuroimage*, *55*(4), 1816-1824.

Goodale, E., & Kotagama, S. W. (2006). Context-dependent vocal mimicry in a passerine bird. *Proceedings of the Royal Society of London B: Biological Sciences*, *273*(1588), 875-880.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, *15*(1), 20-25.

Goto, Y., & Grace, A. A. (2005). Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nature neuroscience*, *8*(6), 805.

Goucha, T., & Friederici, A. D. (2015). The language skeleton after dissecting meaning: a functional segregation within Broca's Area. *Neuroimage*, *114*, 294-302.

Goucha, T., Zaccarella, E., & Friederici, A. D. (2017). A revival of the Homo loquens as a builder of labeled structures: neurocognitive considerations. *Neuroscience & Biobehavioral Reviews*.

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and cognitive processes*, *25*(2), 149-188.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, *100*(1), 253-258.

Grimsley, J. M., Monaghan, J. J., & Wenstrup, J. J. (2011). Development of social vocalizations in mice. *PloS one*, *6*(3), e17460.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive science*, *29*(2), 261-290.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in cognitive sciences*, *9*(9), 416-423.

Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in Psychology*, *4*.

Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Current opinion in neurobiology*, *28*, 136-141.

Hagoort, P., & Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual Review of Neuroscience*, *37*, 347-362.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and cognitive processes*, *8*(4), 439-483.

Hale, J. (2016). Information-theoretical Complexity Metrics. *Language and Linguistics Compass*, *10*(9), 397-412.

Hamao, S., & Eda-Fujiwara, H. (2004). Vocal mimicry by the Black-browed Reed Warbler Acrocephalus bistrigiceps: objective identification of mimetic sounds. *Ibis*, *146*(1), 61-68.

Hammerschmidt, K., Schreiweis, C., Minge, C., Pääbo, S., Fischer, J., & Enard, W. (2015). A humanized version of Foxp2 does not affect ultrasonic vocalization in adult mice. *Genes, Brain and Behavior*, *14*(8), 583-590.

Hardus, M. E., Lameira, A. R., Van Schaik, C. P., & Wich, S. A. (2009). Tool use in wild orang-utans modifies sound production: a functionally deceptive innovation? *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20091027.

Hare, B., Wobber, V., & Wrangham, R. (2012). The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. *Animal Behaviour*, *83*(3), 573-585.

Harmony, T. (2013). The functional significance of delta oscillations in cognitive processing. *Frontiers in integrative neuroscience*, *7*.

Hartwigsen, G., Henseler, I., Stockert, A., Wawrzyniak, M., Wendt, C., Klingbeil, J., ... & Saur, D. (2017). Integration demands modulate effective connectivity in a fronto-temporal network for contextual sentence integration. *NeuroImage*, *147*, 812-824.

Harvey, D. Y., Wei, T., Ellmore, T. M., Hamilton, A. C., & Schnur, T. T. (2013). Neuropsychological evidence for the functional role of the uncinate fasciculus in semantic control. *Neuropsychologia*, *51*(5), 789-801.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569-1579.

Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., ... & Lewontin, R. C. (2014). The mystery of language evolution. *Frontiers in psychology*, *5*.

Healy, S. D., & Rowe, C. (2007). A critique of comparative studies of brain size. *Proceedings of the Royal Society of London B: Biological Sciences*, *274*(1609), 453-464.

Heinrich, B. (1995). An experimental investigation of insight in common ravens (Corvus corax). *The Auk*, 994-1003.

Helmholtz, H. (1860/1962). *Handbuch der physiologischen optik*. (English trans., Southall JPC, Ed.) Vol. 3. Dover, New York.

Herculano-Houzel, S. (2012). The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proceedings of the National Academy of Sciences*, *109*(Supplement 1), 10661-10668.

Herculano-Houzel, S. (2017). Numbers of neurons as biological correlates of cognitive capability. *Current Opinion in Behavioral Sciences*, *16*, 1-7.

Herculano-Houzel, S., Kaas, J. H., & de Oliveira-Souza, R. (2015). Corticalization of motor control in humans is a consequence of brain scaling in primate evolution. *Journal of Comparative Neurology*.

Herculano-Houzel, S., Manger, P. R., & Kaas, J. H. (2014). Brain scaling in mammalian evolution as a consequence of concerted and mosaic changes in numbers of neurons and average neuronal cell size. *Frontiers in neuroanatomy*, *8*.

Hessler, N. A., & Doupe, A. J. (1999). Social context modulates singing-related neural activity in the songbird forebrain. *Nature neuroscience*, *2*(3), 209-211.

Hickok, G. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of communication disorders*, *45*(6), 393-402.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1), 67-99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature reviews. Neuroscience*, *8*(5), 393.

Hill, E. L. (2001). Non-specific nature of specific language impairment: a review of the literature with regard to concomitant motor impairments. *International Journal of Language & Communication Disorders*, *36*(2), 149-171.

Hindmarsh, A. M. (1986). The functional significance of vocal mimicry in song. *Behaviour*, *99*(1), 87-100.

Hosino, T., & Okanoya, K. (2000). Lesion of a higher-order song nucleus disrupts phrase level complexity in Bengalese finches. *Neuroreport*, *11*(10), 2091-2095.

Howard, D., & Patterson, K. E. (1992). *The Pyramids and Palm Trees Test: A test of semantic access from words and pictures*. Thames Valley Test Company.

Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2007). Time course of semantic processes during sentence comprehension: an fMRI study. *Neuroimage*, *36*(3), 924-932.

Humphries, C., Love, T., Swinney, D., & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Human brain mapping*, *26*(2), 128-138.

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., ... & Handwerker, D. A. (2013). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, *80*, 360-378.

Jackendoff, R. (2002). English particle constructions, the lexicon, and the autonomy of syntax. *Verb-particle explorations*, 67-94.

Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain research*, *1146*, 2-22.

Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, *1016*(1), 749-777.

Jarvis, E. D. (2006). Selection for and against vocal learning in birds and mammals. *Ornithological Science*, *5*(1), 5-14.

Jarvis, E. D. (2007). Neural systems for vocal learning in birds and humans: a synopsis. *Journal of Ornithology*, *148*(1), 35-44.

Jarvis, E. D. (2017). Birdsong and the Neurobiology of Language. *Annual Review of Linguistics*, *3*(1).

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., ... & Suh, A. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, *346*(6215), 1320-1331.

Jarvis, E. D., Scharff, C., Grossman, M. R., Ramos, J. A., & Nottebohm, F. (1998). For whom the bird sings: context-dependent gene expression. *Neuron*, *21*(4), 775-788.

Jaynes, E. T. (1988). How does the brain do plausible reasoning? In *Maximum-entropy and Bayesian methods in science and engineering* (pp. 1-24). Springer Netherlands.

Jeon, H. A., & Friederici, A. D. (2013). Two principles of organization in the prefrontal cortex are cognitive hierarchy and degree of automaticity. *Nature Communications*, *4*, 2041.

Jeon, H. A., & Friederici, A. D. (2015). Degree of automaticity and the prefrontal cortex. *Trends in cognitive sciences*, *19*(5), 244-250.

Jeon, H. A., Anwander, A., & Friederici, A. D. (2014). Functional network mirrored in the prefrontal cortex, caudate nucleus, and thalamus: high-resolution functional imaging and structural connectivity. *Journal of Neuroscience*, *34*(28), 9202-9212.

Johnson, M. H. (2011). Interactive specialization: a domain-general framework for human functional brain development? *Developmental cognitive neuroscience*, *1*(1), 7-21.

Joshi, A. K. (1985). Tree adjoining grammars: How much context-sensitivity is required to provide reasonable structural descriptions? In D. R. Dowty, L. Karttunen & A. M. Zwicky, (Eds.), *Natural Language Parsing*. Cambridge. Cambridge University Press.

Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behavioral and Brain Sciences*, *30*(2), 135-154.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and cognitive processes*, *15*(2), 159-201.

Kaplan, G. (1999). Song Structure and Function of Mimicry in the Australian Magpie (Gymnorhina tibicen): Compared to Lyrebird (Menura ssp.). *International Journal of Comparative Psychology*, *12*(4).

Karmiloff-Smith, A. (2015). An alternative to domain-general or domain-specific frameworks for theorizing about human evolution and ontogenesis. *AIMS neuroscience*, *2*(2), 91.

Kaskan, P. M., Franco, E. C. S., Yamada, E. S., de Lima Silveira, L. C., Darlington, R. B., & Finlay, B. L. (2005). Peripheral variability and central constancy in mammalian visual system evolution. *Proceedings of the Royal Society of London B: Biological Sciences*, *272*(1558), 91-100.

Kelley, L. A., & Healy, S. D. (2011). Vocal mimicry. *Current Biology*, *21*(1), R9-R10.

Kelley, L. A., Coe, R. L., Madden, J. R., & Healy, S. D. (2008). Vocal mimicry in songbirds. *Animal behaviour*, *76*(3), 521-528.

Khateb, A., Pegna, A. J., Landis, T., Mouthon, M. S., & Annoni, J. M. (2010). On the origin of the N400 effects: an ERP waveform and source localization analysis in three matching tasks. *Brain topography*, *23*(3), 311-320.

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS computational biology*, *4*(11), e1000209.

Kilner, J. M., & Frith, C. D. (2008). Action observation: inferring intentions without mirror neurons. *Current Biology*, *18*(1), R32-R33.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive processing*, *8*(3), 159-166.

Kim, A., & Lai, V. (2012). Rapid interactions between lexical semantic and word form analysis during word recognition in context: Evidence from ERPs. *Journal of Cognitive Neuroscience*, *24*(5), 1104-1112.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of memory and Language*, *52*(2), 205-225.

Kim, A., & Sikos, L. (2011). Conflict and surrender during sentence processing: An ERP study of syntax-semantics interaction. *Brain and Language*, *118*(1), 15-22.

Kinzler, K. D., & Spelke E.S. (2007). Core systems in human cognition. *Progress in brain research* 164: 257-264.

Kirchhof, J., & Hammerschmidt, K. (2006). Functionally referential alarm calls in tamarins (Saguinus fuscicollis and Saguinus mystax)–evidence from playback experiments. *Ethology*, *112*(4), 346-354.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology*, *55*(4), 352.

Kitzbichler, M. G., Henson, R. N., Smith, M. L., Nathan, P. J., & Bullmore, E. T. (2011). Cognitive effort drives workspace configuration of human brain functional networks. *The Journal of Neuroscience*, *31*(22), 8259-8270.

Kitzbichler, M. G., Smith, M. L., Christensen, S. R., & Bullmore, E. (2009). Broadband criticality of human brain network synchronization. *PLoS Comput Biol*, *5*(3), e1000314.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*(5648), 1181-1185.

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in cognitive sciences*, *11*(6), 229-235.

Konishi, M., McLaren, D. G., Engen, H., & Smallwood, J. (2015). Shaped by the past: the default mode network supports cognition that is independent of immediate perceptual input. *PLoS One*, *10*(6), e0132209.

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in cognitive sciences*, *17*(1), 26-49.

Kukona, A., Fang, S. Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, *119*(1), 23-42.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain research*, *1146*, 23-49.

Kuperberg, G. R. (2008). Electroencephalography, event-related potentials, and magnetoencephalography. *Essentials of neuroimaging for clinical practice*, 117-127.

Kuperberg, G. R. (2013). The proactive comprehender: What event-related potentials tell us about the dynamics of reading comprehension. *Unraveling the behavioral, neurobiological, and genetic components of reading comprehension*, 176-192.

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, cognition and neuroscience*, *31*(5), 602-616.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, cognition and neuroscience*, *31*(1), 32-59.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience*, *15*(2), 272-293.

Kuperberg, G. R., McGuire, P. K., Bullmore, E. T., Brammer, M. J., Rabe-Hesketh, S., Wright, I. C., ... & David, A. S. (2000). Common and distinct neural substrates for pragmatic, semantic, and syntactic processing of spoken sentences: an fMRI study. *Journal of Cognitive Neuroscience*, *12*(2), 321-341.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621-647.

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological psychology*, *11*(2), 99-116.

Kutas, M., Federmeier, K. D., & Urbach, T. (2014). The “negatives” and “positives” of prediction in language. *The Cognitive Neurosciences V*, 649-56.

Kuypers H. G. (1958) Corticobulbar connexions to the pons and lower brain-stem in man. *Brain*, 81, 364–388.

Lameira, A. R., Hardus, M. E., Mielke, A., Wich, S. A., & Shumaker, R. W. (2016). Vocal fold control beyond the species-specific repertoire in an orang-utan. *Scientific reports*, *6*.

Langmore, N. E. (2000). Why female birds sing. *Animal Signals. Tapir Academic Press, Trondheim*, 317-327.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (ed.), *Cerebral mechanisms in behavior*. New York: Wiley.

Lau, E. F., & Nguyen, E. (2015). The role of temporal predictability in semantic expectation: An MEG investigation. *cortex, 68,* 8-19.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature reviews. Neuroscience*, *9*(12), 920.

Lawrence, D. G., & Kuypers, H. G. (1968). The functional organization of the motor system in the monkey 1:II. The effects of lesions of the descending brain-stem pathways. *Brain,* 91(1), 15-36.

Laycock, R., Crewther, S. G., & Crewther, D. P. (2007). A role for the 'magnocellular advantage'in visual impairments in neurodevelopmental and psychiatric disorders. *Neuroscience & Biobehavioral Reviews*, *31*(3), 363-376.

Leech, R., & Sharp, D. J. (2014). The role of the posterior cingulate cortex in cognition and disease. *Brain*, *137*(1), 12-32.

Lefebvre, L. (2013). Brains, innovations, tools and cultural transmission in birds, non-human primates, and fossil hominins. *Frontiers in human neuroscience*, *7*.

Lelekov-Boissard, T., & Dominey, P. F. (2002). Human brain potentials reveal similar processing of non-linguistic abstract structure and linguistic syntactic structure. *Neurophysiologie Clinique/Clinical Neurophysiology, 32*(1), 72-84.

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Phil. Trans. R. Soc. B*, *369*(1651), 20130302.

Lill, A. & Boesman, P. (2017). *Superb Lyrebird (Menura novaehollandiae).* In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A. & de Juana, E. (eds.). *Handbook of the Birds of the World Alive*. Lynx Edicions, Barcelona.

Lisman, J. E., & Jensen, O. (2013). The theta-gamma neural code. *Neuron*, *77*(6), 1002-1016.

Liu, W. C., Wada, K., Jarvis, E. D., & Nottebohm, F. (2013). Rudimentary substrates for vocal learning in a suboscine. *Nature communications*, *4*.

Loepelt, J., Shaw, R. C., & Burns, K. C. (2016, June). Can you teach an old parrot new tricks? Cognitive development in wild kaka (Nestor meridionalis). In *Proc. R. Soc. B* (Vol. 283, No. 1832, p. 20153056). The Royal Society.

Ludwig, E., & Klinger, J. (1956). *Atlas Cerebris Humani: The Inner Structure of the Brain.* Basel, Switzerland: Little, Brown and Company.

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science, 24*(4), 279-284.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in psychology, 4*.

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., ... & Boogert, N. J. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences*, *111*(20), E2140-E2148.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature neuroscience*, *4*(5), 540.

Makuuchi, M., & Friederici, A. D. (2013). Hierarchical functional connectivity between the core language system and the working memory system. *Cortex*, *49*(9), 2416-2423.

Makuuchi, M., Bahlmann, J., Anwander, A., & Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences*, *106*(20), 8362-8367.

Marcus, G. F. (2006). Cognitive architecture and descent with modification. *Cognition*, *101*(2), 443-465.

Marler, P. (1977). The structure of animal communication sounds. *Recognition of complex acoustic signals*, 17-35.

Marler, P. (2004). Bird calls: their potential for behavioral neurobiology. *Annals of the New York Academy of Sciences*, *1016*(1), 31-44.

Marslen-Wilson, W. D., Bozic, M., & Tyler, L. K. (2014). Morphological systems in their neurobiological contexts. *The Cognitive Neurosciences, 5th ed. MIT Press, Cambridge, MA.*

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science*, *315*(5810), 393-395.

Matchin, W., & Hickok, G. (2016). 'Syntactic Perturbation'During Production Activates the Right IFG, but not Broca's Area or the ATL. *Frontiers in psychology, 7*.

Matchin, W., Hammerly, C., & Lau, E. (2017). The role of the IFG and pSTS in syntactic prediction: Evidence from a parametric study of hierarchical structure in fMRI. *cortex, 88,* 106-123.

Matchin, W., Sprouse, J., & Hickok, G. (2014). A structural distance effect for backward anaphora in Broca's area: An fMRI study. *Brain and language, 138,* 1-11.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in cognitive sciences*, *10*(8), 363-369.

McDonough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., & Lannon, R. (2011). An image is worth a thousand words: Why nouns tend to dominate verbs in early word learning. *Developmental science*, *14*(2), 181-189.

McKiernan, K. A., D'Angelo, B. R., Kaufman, J. N., & Binder, J. R. (2006). Interrupting the "stream of consciousness": an fMRI investigation. *Neuroimage*, *29*(4), 1185-1191.

McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning*, *60*(s2), 123-150.

Mello, C. V. (2014). The zebra finch, Taeniopygia guttata: An avian model for investigating the neurobiological basis of vocal learning. *Cold Spring Harbor Protocols, 2014*(12), pdb-emo084574.

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, *214*(5-6), 655-667.

Mercado, E., Mantell, J. T., & Pfordresher, P. Q. (2014). Imitating sounds: A cognitive approach to understanding vocal imitation. *Comparative Cognition & Behavior Reviews*, *9*.

Mesulam, M. M. (1998). From sensation to cognition. *Brain*, *121*(6), 1013-1052.

Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., & Friederici, A. D. (2016). Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cerebral Cortex,* 1-10.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, *24*(1), 167-202.

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural brain research*, *6*(1), 57-77.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2009). Statistical learning of nonadjacencies predicts on-line processing of long-distance dependencies in natural language. In *Proceedings of the Cognitive Science Society* (pp. 177-182).

Miyagawa, S. (2016). Integration hypothesis: A parallel model of language development in evolution. *Evolution of the brain, cognition, and emotion in vertebrates. Berlin: Springer*.

Miyagawa, S., Berwick, R. C., & Okanoya, K. (2013). The emergence of hierarchical structure in human language. *Frontiers in psychology*, *4*.

Miyagawa, S., Ojima, S., Berwick, R. C., & Okanoya, K. (2014). The integration hypothesis of human language evolution and the nature of contemporary languages. *Frontiers in psychology*, *5*.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive psychology*, *41*(1), 49-100.

Mohr, H., Wolfensteller, U., Betzel, R. F., Mišić, B., Sporns, O., Richiardi, J., & Ruge, H. (2016). Integration and segregation of large-scale brain networks during short-term task automatization. *Nature communications*, *7*.

Moore, B. R. (2004). The evolution of learning. *Biological Reviews*, *79*(2), 301-335.

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological cybernetics*, *66*(3), 241-251.

Nagahama, Y., Okada, T., Katsumi, Y., Hayashi, T., Yamauchi, H., Sawamoto, N., ... & Fukuyama, H. (1999). Transient neural activity in the medial superior frontal gyrus and precuneus time locked with attention shift between object features. *Neuroimage*, *10*(2), 193-199.

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience*, *22*(12), 2886-2898.

Nee, D. E., & D'Esposito, M. (2016). The hierarchical organization of the lateral prefrontal cortex. *Elife*, *5*, e12112.

Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., ... & Dehaene, S. (2017a). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*, 201701590.

Nelson, M., Dehaene, S., Pallier, C., & Hale, J. (2017b). Entropy Reduction correlates with temporal lobe activity. In *Proceedings of the 7th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2017)* (pp. 1-10).

Neubert, F. X., Mars, R. B., Thomas, A. G., Sallet, J., & Rushworth, M. F. (2014). Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron*, *81*(3), 700-713.

Neufeld, C., Kramer, S. E., Lapinskaya, N., Heffner, C. C., Malko, A., & Lau, E. F. (2016). The Electrophysiology of Basic Phrase Building. *PloS one*, *11*(10), e0158446.

Nicenboim, B., Logačev, P., Gattei, C., & Vasishth, S. (2016). When high-capacity readers slow down and low-capacity readers speed up: Working memory and locality effects. *Frontiers in psychology*, *7*.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus--norepinephrine system. *Psychological bulletin*, *131*(4), 510.

Nottebohm, F. (1991). Reassessing the mechanisms and origins of vocal learning in birds. *Trends in neurosciences*, *14*(5), 206-211.

Nottebohm, F., & Arnold, A. P. (1976). Sexual dimorphism in vocal control areas of the songbird brain. *Science*, *194*(4261), 211-213.

Nottebohm, F., Stokes, T. M., & Leonard, C. M. (1976). Central control of song in the canary, Serinus canarius. *Journal of Comparative Neurology*, *165*(4), 457-486.

Núñez-Peña, M. I., & Honrubia-Serrano, M. L. (2004). P600 related to rule violation in an arithmetic task. *Cognitive Brain Research*, *18*(2), 130-141.

Oberauer, K. (2013). The focus of attention in working memory—from metaphors to mechanisms. *Frontiers in human neuroscience*, *7*.

Odom, K. J., Hall, M. L., Riebel, K., Omland, K. E., & Langmore, N. E. (2014). Female song is widespread and ancestral in songbirds. *Nature Communications*, *5*, 3379.

Okanoya, K. (2004). Song syntax in Bengalese finches: proximate and ultimate analyses. *Advances in the Study of Behavior*, *34*, 297-346.

Okanoya, K. (2012). Behavioural factors governing song complexity in Bengalese finches. *International Journal of Comparative Psychology*, *25*(1).

Okanoya, K. (2015). Evolution of song complexity in Bengalese finches could mirror the emergence of human language. *Journal of Ornithology*, *156*(1), 65-72.

Olson, I. R., Plotzker, A., & Ezzyat, Y. (2007). The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain*, *130*(7), 1718-1731.

Olson, I. R., Von Der Heide, R. J., Alm, K. H., & Vyas, G. (2015). Development of the uncinate fasciculus: Implications for theory and developmental disorders. *Developmental cognitive neuroscience*, *14*, 50-61.

Onnis, L., Christiansen, M. H., Chater, N., & Gómez, R. (2003, January). Reduction of uncertainty in human sequential learning: Evidence from artificial grammar learning. In *Proceedings of the 25th annual conference of the cognitive science society* (pp. 886-891). Malwah, NJ: Lawrence Erlbaum Associates.

Onnis, L., Destrebecqz, A., Christiansen, M. H., Chater, N., & Cleeremans, A. (2015). Implicit learning of non-adjacent dependencies. *Implicit and Explicit Learning of Languages, 48*, 213.

Opitz, B., & Friederici, A. D. (2007). Neural basis of processing sequential and hierarchical syntactic structures. *Human brain mapping, 28*(7), 585-592.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice Learners, Longitudinal Designs, and Event-Related Potentials: A Means for Exploring the Neurocognition of Second Language Processing. *Language Learning, 56*(s1), 199-230.

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys use affixation to alter call meaning. *PloS one, 4*(11), e7808.

Paczynski, M., & Kuperberg, G. R. (2011). Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb-argument processing. *Language and cognitive processes, 26*(9), 1402-1456.

Pallier, C., Devauchelle, A. D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences, 108*(6), 2522-2527.

Papagno, C., Miracapillo, C., Casarotti, A., Romero Lauro, L. J., Castellano, A., Falini, A., ... & Bello, L. (2010). What is the role of the uncinate fasciculus? Surgical removal and proper name retrieval. *Brain, 134*(2), 405-414.

Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: selected studies of the event-related potential. *International journal of medical sciences, 2*(4), 147.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature reviews. Neuroscience, 8*(12), 976.

Peelle, J. E., Johnsrude, I. S., & Davis, M. H. (2010). Hierarchical processing for speech in human auditory cortex and beyond. *Frontiers in human neuroscience, 4*.

Pepperberg, I. M. (1981). Functional vocalizations by an African Grey parrot (Psittacus erithacus). *Ethology, 55*(2), 139-160.

Pepperberg, I. M. (1990). Cognition in an African gray parrot (Psittacus erithacus): Further evidence for comprehension of categories and labels. *Journal of Comparative Psychology*, *104*(1), 41.

Pepperberg, I. M. (2006). Cognitive and communicative abilities of Grey parrots. *Applied Animal Behaviour Science*, *100*(1), 77-86.

Pepperberg, I. M. (2010). Vocal learning in Grey parrots: A brief review of perception, production, and cross-species comparisons. *Brain and language*, *115*(1), 81-91.

Perani, D., Saccuman, M. C., Scifo, P., Anwander, A., Spada, D., Baldoli, C., ... & Friederici, A. D. (2011). Neural language networks at birth. *Proceedings of the National Academy of Sciences*, *108*(38), 16056-16061.

Petkov, C. I., & Jarvis, E. D. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Frontiers in evolutionary neuroscience*, *4*.

Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, *134*, 17-35.

Pfaff, J. A., Zanette, L., MacDougall-Shackleton, S. A., & MacDougall-Shackleton, E. A. (2007). Song repertoire size varies with HVC volume and is indicative of male quality in song sparrows (Melospiza melodia). *Proceedings of the Royal Society of London B: Biological Sciences*, *274*(1621), 2035-2040.

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., ... & Mountcastle, J. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*, *346*(6215), 1256846.

Phillips, C., & Lewis, S. (2013). Derivational order in syntax: Evidence and architectural consequences. *Studies in Linguistics*, *6*, 11-47.

Piai, V., Anderson, K. L., Lin, J. J., Dewar, C., Parvizi, J., Dronkers, N. F., & Knight, R. T. (2016). Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. *Proceedings of the National Academy of Sciences*, 201603312.

Piccoli, T., Valente, G., Linden, D. E., Re, M., Esposito, F., Sack, A. T., & Di Salle, F. (2015). The default mode network and the working memory network are not anti-correlated during all phases of a working memory task. *PloS one, 10*(4), e0123354.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in cognitive sciences*, *11*(3), 105-110.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329-347.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition*, *95*(2), 201-236.

Pisanski, K., Cartei, V., McGettigan, C., Raine, J., & Reby, D. (2016). Voice modulation: A window into the origins of human vocal control? *Trends in cognitive sciences*, *20*(4), 304-318.

Plotnik, J. M., & Clayton, N. S. (2015). Convergent cognitive evolution across animal taxa: comparisons of chimpanzees, corvids, and elephants. *The conceptual mind: new directions in the study of concepts. MIT Press, Cambridge*, 29-55.

Poeppel, D. (2012). The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language. *Cognitive neuropsychology*, *29*(1-2), 34-55.

Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Current opinion in neurobiology*, *28*, 142-149.

Poeppel, D., & Embick, D. (2005). Defining the relation between linguistics and neuroscience. *Twenty-first century psycholinguistics: Four cornerstones*, 103-118.

Poeppel, D., Emmorey, K., Hickok, G., & Pylkkänen, L. (2012). Towards a new neurobiology of language. *Journal of Neuroscience*, *32*(41), 14125-14131.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128-2148.

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). Animal behaviour: elephants are capable of vocal learning. *Nature*, *434*(7032), 455-456.

Poortman, E. B., & Pylkkänen, L. (2016). Adjective conjunction as a window into the LATL’s contribution to conceptual combination. *Brain and language*, *160*, 50-60.

Popa, D., Popescu, A. T., & Paré, D. (2009). Contrasting activity profile of two distributed cortical networks as a function of attentional demands. *Journal of Neuroscience*, *29*(4), 1191-1201.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., ... & Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron*, *72*(4), 665-678.

Price, T., Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K., & Fischer, J. (2015). Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Scientific reports*, *5*.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature reviews. Neuroscience*, *11*(5), 351.

Pylkkänen, L., Bemis, D. K., & Elorrieta, E. B. (2014b). Building phrases in language production: An MEG study of simple composition. *Cognition*, *133*(2), 371-384.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, *80*(1), 127-158.

Raichle, M. E. (2015). The brain's default mode network. *Annual review of neuroscience*, *38*, 433-447.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676-682.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1).

Rao, R. P., & Sejnowski, T. J. (2002). 16 Predictive Coding, Cortical Feedback, and Spike-Timing Dependent Plasticity. *Probabilistic models of the brain*, 297.

Rauschecker, J. P. (2012). Ventral and dorsal streams in the evolution of speech and language. *Frontiers in evolutionary neuroscience*, *4*.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proceedings of the National Academy of Sciences*, *97*(22), 11800-11806.

Redish, A. D. (1999). *Beyond the cognitive map: from place cells to episodic memory*. MIT Press.

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn't it ironic? An electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience*, *23*(2), 277-293.

Reid, J. M., Arcese, P., Cassidy, A. L., Hiebert, S. M., Smith, J. N., Stoddard, P. K., ... & Keller, L. F. (2004). Song repertoire size predicts initial mating success in male song sparrows, Melospiza melodia. *Animal Behaviour*, *68*(5), 1055-1063.

Richards, D. G., Wolz, J. P., & Herman, L. M. (1984). Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, Tursiops truncatus.. *Journal of Comparative Psychology*, *98*(1), 10.

Riebel, K. (2003). The “mute” sex revisited: vocal production and perception learning in female songbirds. *Advances in the Study of Behavior*, *33*, 49-86.

Rogalsky, C., & Hickok, G. (2008). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cerebral Cortex*, *19*(4), 786-796.

Rogalsky, C., Poppa, T., Chen, K. H., Anderson, S. W., Damasio, H., Love, T., & Hickok, G. (2015). Speech repetition as a window on the neurobiology of auditory–motor integration for speech: A voxel-based lesion symptom mapping study. *Neuropsychologia*, *71*, 18-27.

Rolheiser, T., Stamatakis, E. A., & Tyler, L. K. (2011). Dynamic processing in the human language system: synergy between the arcuate fascicle and extreme capsule. *Journal of Neuroscience*, *31*(47), 16949-16957.

Rommers, J., Dickson, D. S., Norton, J. J., Wlotko, E. W., & Federmeier, K. D. (2017). Alpha and theta band dynamics related to sentential constraint and word expectancy. *Language, Cognition and Neuroscience*, *32*(5), 576-589.

Rushworth, M. F., & Behrens, T. E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature neuroscience*, *11*(4), 389.

Rusli, Y. A., & Montgomery, J. W. (2017). Children's Comprehension of Object Relative Sentences: It's Extant Language Knowledge That Matters, Not Domain-General Working Memory. *Journal of Speech, Language, and Hearing Research*, 1-14.

Sadaghiani, S., & D'Esposito, M. (2014). Functional characterization of the cingulo-opercular network in the maintenance of tonic alertness. *Cerebral Cortex*, *25*(9), 2763-2773.

Sale, M. V., Reid, L. B., Cocchi, L., Pagnozzi, A. M., Rose, S. E., & Mattingley, J. B. (2017). Brain changes following four weeks of unimanual motor training: Evidence from behavior, neural stimulation, cortical thickness, and functional MRI. *Human Brain Mapping*, *38*(9), 4773-4787.

Sanford, A. J., & Graesser, A. C. (2006). Shallow processing and underspecification. *Discourse Processes*, *42*(2), 99-108.

Santi, A., & Grodzinsky, Y. (2007). Working memory and syntax interact in Broca's area. *Neuroimage*, *37*(1), 8-17.

Santi, A., & Grodzinsky, Y. (2010). fMRI adaptation dissociates syntactic complexity dimensions. *Neuroimage*, *51*(4), 1285-1293.

Santi, A., & Grodzinsky, Y. (2012). Broca's area and sentence comprehension: A relationship parasitic on dependency, displacement or predictability? *Neuropsychologia*, *50*(5), 821-832.

Santi, A., Friederici, A. D., Makuuchi, M., & Grodzinsky, Y. (2015). An fMRI study dissociating distance measures computed by Broca's area in movement processing: clause boundary vs. identity. *Frontiers in psychology*, *6*.

Sapolsky, R. M. (2006). A natural history of peace. *Foreign Affairs*, 104-120.

Sassenhagen, J., & Bornkessel-Schlesewsky, I. (2015). The P600 as a correlate of ventral attention network reorientation. *Cortex*, *66*, A3-A20.

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M. S., ... & Huber, W. (2008). Ventral and dorsal pathways for language. *Proceedings of the national academy of Sciences*, *105*(46), 18035-18040.

Scharff, C., & Nottebohm, F. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *Journal of Neuroscience*, *11*(9), 2896-2913.

Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2009). When semantic P600s turn into N400s. *Brain Talk*, 75.

Schreiweis, C., Bornschein, U., Burguière, E., Kerimoglu, C., Schreiter, S., Dannemann, M., ... & Groszer, M. (2014). Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proceedings of the National Academy of Sciences*, *111*(39), 14253-14258.

Seed, A., Emery, N., & Clayton, N. (2009). Intelligence in corvids and apes: a case of convergent evolution? *Ethology*, *115*(5), 401-420.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., ... & Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, *27*(9), 2349-2356.

Semendeferi, K., Armstrong, E., Schleicher, A., Zilles, K., & Van Hoesen, G. W. (2001). Prefrontal cortex in humans and apes: a comparative study of area 10. *American journal of physical anthropology*, *114*(3), 224-241.

Sestieri, C., Corbetta, M., Spadone, S., Romani, G. L., & Shulman, G. L. (2014). Domain-general signals in the cingulo-opercular network for visuospatial attention and episodic memory. *Journal of cognitive neuroscience*, *26*(3), 551-568.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Animal Behaviour*, *28*(4), 1070-1094.

Shanahan, M. (2010). Metastable chimera states in community-structured oscillator networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *20*(1), 013108.

Shannon, C. E. (1948). A mathematical theory of communication, Part I, Part II. *Bell Syst. Tech. J.*, *27*, 623-656.

Shapiro, K. A., Mottaghy, F. M., Schiller, N. O., Poeppel, T. D., Flüß, M. O., Müller, H. W., ... & Krause, B. J. (2005). Dissociating neural correlates for nouns and verbs. *Neuroimage*, *24*(4), 1058-1067.

Shen, W., Fiori-Duharcourt, N., & Isel, F. (2016). Functional significance of the semantic P600: evidence from the event-related brain potential source localization. *NeuroReport*, *27*(7), 548-558.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217-240.

Sheppard, J. P., Wang, J. P., & Wong, P. C. (2012). Large-scale cortical network properties predict future sound-to-word learning success. *Journal of cognitive neuroscience*, *24*(5), 1087-1103.

Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E., & Petersen, S. E. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *Journal of cognitive neuroscience*, *9*(5), 648-663.

Simons, J. S., Peers, P. V., Hwang, D. Y., Ally, B. A., Fletcher, P. C., & Budson, A. E. (2008). Is the parietal lobe necessary for recollection in humans? *Neuropsychologia*, *46*(4), 1185-1191.

Simonyan, K., & Jürgens, U. (2003). Efferent subcortical projections of the laryngeal motorcortex in the rhesus monkey. *Brain research*, *974*(1), 43-59.

Skeide, M. A., & Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nature Reviews Neuroscience*, *17*(5), 323-332.

Skeide, M. A., Brauer, J., & Friederici, A. D. (2015). Brain functional and structural predictors of language performance. *Cerebral Cortex*, *26*(5), 2127-2139.

Skipper, L. M., Ross, L. A., & Olson, I. R. (2011). Sensory and semantic category subdivisions within the anterior temporal lobes. *Neuropsychologia*, *49*(12), 3419-3429.

Smallwood, J. (2013). Distinguishing how from why the mind wanders: a process–occurrence framework for self-generated mental activity. *Psychological bulletin*, *139*(3), 519.

Smallwood, J., Brown, K., Baird, B., & Schooler, J. W. (2012). Cooperation between the default mode network and the frontal–parietal network in the production of an internal train of thought. *Brain research*, *1428*, 60-70.

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1-B11.

Sonuga-Barke, E. J., & Castellanos, F. X. (2007). Spontaneous attentional fluctuations in impaired states and pathological conditions: a neurobiological hypothesis. *Neuroscience & Biobehavioral Reviews*, *31*(7), 977-986.

Spelke, E. S. (1990). Principles of object perception. *Cognitive science* 14.1: 29-56.

Spelke, E. S. (2003). What makes us smart? Core knowledge and natural language. *Language in mind: Advances in the study of language and thought*, 277-311.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological review*, *99*(4), 605.

Spreng, R. N., & Grady, C. L. (2010). Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network. *Journal of cognitive neuroscience*, *22*(6), 1112-1123.

Sridharan, D., Levitin, D. J., & Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proceedings of the National Academy of Sciences*, *105*(34), 12569-12574.

Steele, V. R., Bernat, E. M., van den Broek, P., Collins, P. F., Patrick, C. J., & Marsolek, C. J. (2013). Separable processes before, during, and after the N400 elicited by previously inferred and new information: Evidence from time–frequency decompositions. *Brain research*, *1492*, 92-107.

Stevens, J. R. (2014). Evolutionary pressures on primate intertemporal choice. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*(1786), 20140499.

Stoeger, A. S., Mietchen, D., Oh, S., de Silva, S., Herbst, C. T., Kwon, S., & Fitch, W. T. (2012). An Asian elephant imitates human speech. *Current Biology*, *22*(22), 2144-2148.

Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, *78*(2), 364-375.

Stout, D., Hecht, E., Khreisheh, N., Bradley, B., & Chaminade, T. (2015). Cognitive demands of Lower Paleolithic toolmaking. *PLoS One*, *10*(4), e0121804.

Stowe, L. A., Broere, C. A., Paans, A. M., Wijers, A. A., Mulder, G., Vaalburg, W., & Zwarts, F. (1998). Localizing components of a complex task: sentence processing and working memory. *Neuroreport*, *9*(13), 2995-2999.

Strijkers, K., & Costa, A. (2016). The cortical dynamics of speaking: Present shortcomings and future avenues. *Language, Cognition and Neuroscience*, *31*(4), 484-503.

Stroud, C., & Phillips, C. (2012). Examining the evidence for an independent semantic analyzer: An ERP study in Spanish. *Brain and language*, *120*(2), 108-126.

Struhsaker, T. T. (1967). Auditory communication among vervet monkeys (Cercopithecus aethiops). *Social communication among primates*, 281-324.

Suh, A., Paus, M., Kiefmann, M., Churakov, G., Franke, F. A., Brosius, J., ... & Schmitz, J. (2011). Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nature communications*, *2*, 443.

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2012). Language-related ERP components. *Oxford handbook of event-related potential components*, 397-440.

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology-Learning Memory and Cognition*, *30*(2), 431-449.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, *50*(4), 355-370.

Tanner, D. (2013). Individual differences and streams of processing. *Linguistic Approaches to Bilingualism*, *3*(3), 350-356.

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*(2), 367-382.

Tattersall, I. (2010). Human evolution and cognition. *Theory in Biosciences*, *129*(2-3), 193-201.

Ter Maat, A., Trost, L., Sagunsky, H., Seltmann, S., & Gahr, M. (2014). Zebra finch mates use their forebrain song system in unlearned call communication. *PloS one*, *9*(10), e109334.

Theofanopoulou, C., & Boeckx, C. (2015). Cognitive phylogenies, the Darwinian logic of descent, and the inadequacy of cladistic thinking. *Frontiers in cell and developmental biology*, *3*.

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental science*, *10*(1), 121-125.

Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules* (Vol. 1950). Cambridge, MA: MIT Press.

Traxler, M. J. (2014). Trends in syntactic parsing: Anticipation, Bayesian estimation, and good-enough parsing. *Trends in cognitive sciences*, *18*(11), 605-611.

Traxler, M. J., Long, D. L., Tooley, K. M., Johns, C. L., Zirnstein, M., & Jonathan, E. (2012). Individual differences in eye-movements during reading: working memory and speed-of-processing effects. *Journal of eye movement research*, *5*(1).

Tremblay, P., & Dick, A. S. (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and language*, *162*, 60-71.

Tune, S., Schlesewsky, M., Nagels, A., Small, S. L., & Bornkessel-Schlesewsky, I. (2016). Sentence understanding depends on contextual use of semantic and real world knowledge. *NeuroImage*, *136*, 10-25.

Tyler, L. K., Marslen-Wilson, W. D., Randall, B., Wright, P., Devereux, B. J., Zhuang, J., ... & Stamatakis, E. A. (2011). Left inferior frontal cortex and syntax: function, structure and behaviour in patients with left hemisphere damage. *Brain*, *134*(2), 415-431.

Tyler, L. K., Russell, R., Fadili, J., & Moss, H. E. (2001). The neural representation of nouns and verbs: PET studies. *Brain*, *124*(8), 1619-1634.

Ueno, T., Saito, S., Rogers, T. T., & Ralph, M. A. L. (2011). Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, *72*(2), 385-396.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and cognition*, *4*(2), 105-122.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1), 231-270.

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*(3), 399-433.

Unger, K., Ackerman, L., Chatham, C. H., Amso, D., & Badre, D. (2016). Working memory gating mechanisms explain developmental change in rule-guided behavior. *Cognition*, *155*, 8-22.

Van Den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*(7), 967-985.

Van Den Heuvel, M. P., & Sporns, O. (2011). Rich-club organization of the human connectome. *Journal of Neuroscience*, *31*(44), 15775-15786.

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and language*, *97*(3), 279-293.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176-190.

Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience*, *14*(4), 550-560.

Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., & Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of the National Academy of Sciences*, *92*(3), 930-933.

Vatansever, D., Menon, D. K., Manktelow, A. E., Sahakian, B. J., & Stamatakis, E. A. (2015). Default mode dynamics for global functional integration. *Journal of Neuroscience, 35*(46), 15254-15262.

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience & Biobehavioral Reviews*, *35*(3), 407-426.

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *Journal of neurophysiology, 100*(6), 3328-3342.

Vissiennon, K., Friederici, A. D., Brauer, J., & Wu, C. Y. (2017). Functional organization of the language network in three-and six-year-old children. *Neuropsychologia*, *98*, 24-33.

Vogt, B. A., Vogt, L., & Laureys, S. (2006). Cytology and functionally correlated circuits of human posterior cingulate areas. *Neuroimage*, *29*(2), 452-466.

Von Der Heide, R. J., Skipper, L. M., Klobusicky, E., & Olson, I. R. (2013). Dissecting the uncinate fasciculus: disorders, controversies and a hypothesis. *Brain*, *136*(6), 1692-1707.

Vossel, S., Geng, J. J., & Fink, G. R. (2014). Dorsal and ventral attention systems: distinct neural circuits but collaborative roles. *The Neuroscientist, 20*(2), 150-159.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences, 108*(51), 20754-20759.

Wallentin, M. (2009). Putative sex differences in verbal abilities and language cortex: A critical review. *Brain and language*, *108*(3), 175-183.

Wang, J., Yang, N., Liao, W., Zhang, H., Yan, C. G., Zang, Y. F., & Zuo, X. N. (2015a). Dorsal anterior cingulate cortex in typically developing children: Laterality analysis. *Developmental cognitive neuroscience*, *15*, 117-129.

Wang, L., Uhrig, L., Jarraya, B., & Dehaene, S. (2015b). Representation of numerical and sequential patterns in macaque and human brains. *Current Biology*, *25*(15), 1966-1974.

Wang, R., Chen, C. C., Hara, E., Rivas, M. V., Roulhac, P. L., Howard, J. T., ... & Jarvis, E. D. (2015c). Convergent differential regulation of SLIT-ROBO axon guidance genes in the brains of vocal learners. *Journal of Comparative Neurology*, *523*(6), 892-906.

Waters, G., Caplan, D., Alpert, N., & Stanczak, L. (2003). Individual differences in rCBF correlates of syntactic processing in sentence comprehension: effects of working memory and speed of processing. *NeuroImage*, *19*(1), 101-112.

Weiss-Croft, L. J., & Baldeweg, T. (2015). Maturation of language networks in children: A systematic review of 22years of functional MRI. *NeuroImage*, *123*, 269-281.

Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nature Neuroscience*, *9*(7), 971-978.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive psychology*, *58*(2), 250-271.

Westerlund, M., & Pylkkänen, L. (2014). The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia*, *57*, 59-70.

Wheatcroft, D., & Qvarnström, A. (2015). A blueprint for vocal learning: auditory predispositions from brains to genomes. *Biology letters*, *11*(8), 20150155.

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal fluency: Language or executive function measure? *Applied Neuropsychology: Adult*, *23*(1), 29-34.

Wilbrecht, L., & Nottebohm, F. (2003). Vocal learning in birds and humans. *Developmental Disabilities Research Reviews*, *9*(3), 135-148.

Wilkins, A. S., Wrangham, R. W., & Fitch, W. T. (2014). The “domestication syndrome” in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*, *197*(3), 795-808.

Wilson, B., Marslen-Wilson, W. D., & Petkov, C. I. (2017). Conserved sequence processing in primate frontal cortex. *Trends in neurosciences*.

Wilson, B., Kikuchi, Y., Sun, L., Hunter, D., Dick, F., Smith, K., ... & Petkov, C. I. (2015). Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nature communications*, *6*.

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*(2), 267-279.

Wlotko, E. W., & Federmeier, K. D. (2007). Finding the right word: Hemispheric asymmetries in the use of sentence context information. *Neuropsychologia*, *45*(13), 3001-3014.

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, *62*(1), 356-366.

Wlotko, E. W., & Federmeier, K. D. (2013). Two sides of meaning: The scalp-recorded N400 reflects distinct contributions from the cerebral hemispheres. *Frontiers in psychology*, *4*.

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, *68*, 20-32.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1-34.

Woolgar, A., Hampshire, A., Thompson, R., & Duncan, J. (2011). Adaptive coding of task-relevant information in human frontoparietal cortex. *Journal of Neuroscience*, *31*(41), 14592-14599.

Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-Smith, I., Torralva, T., ... & Duncan, J. (2010). Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, *107*(33), 14899-14902.

Woolsey, T. A., & Van der Loos, H. (1970). The structural organization of layer IV in the somatosensory region (SI) of mouse cerebral cortex: the description of a cortical field composed of discrete cytoarchitectonic units. *Brain research*, *17*(2), 205-242.

Xiao, Y., Friederici, A. D., Margulies, D. S., & Brauer, J. (2016). Development of a selective left-hemispheric fronto-temporal network for processing syntactic complexity in language comprehension. *Neuropsychologia*, *83*, 274-282.

Ye, Z., & Zhou, X. (2008). Involvement of cognitive control in sentence comprehension: Evidence from ERPs. *Brain Research*, *1203*, 103-115.

Ye, Z., & Zhou, X. (2009). Executive control in language processing. *Neuroscience & Biobehavioral Reviews*, *33*(8), 1168-1177.

Yopak, K. E., Lisney, T. J., Darlington, R. B., Collin, S. P., Montgomery, J. C., & Finlay, B. L. (2010). A conserved pattern of brain scaling from sharks to primates. *Proceedings of the National Academy of Sciences*, *107*(29), 12946-12951.

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681-692.

Zaccarella, E., & Friederici, A. D. (2015). Merge in the human brain: A sub-region based functional investigation in the left pars opercularis. *Frontiers in psychology*, *6*.

Zaccarella, E., Meyer, L., Makuuchi, M., & Friederici, A. D. (2015). Building by syntax: the neural basis of minimal linguistic structures. *Cerebral Cortex*, *27*(1), 411-421.

Zaccarella, E., Schell, M., & Friederici, A. D. (2017). Reviewing the functional basis of the syntactic Merge mechanism for language: A coordinate-based activation likelihood estimation meta-analysis. *Neuroscience & Biobehavioral Reviews*, *80*, 646-656.

Zalesky, A., Fornito, A., Cocchi, L., Gollo, L. L., & Breakspear, M. (2014). Time-resolved resting-state brain networks. *Proceedings of the National Academy of Sciences*, *111*(28), 10341-10346.

Zarcone, A., Van Schijndel, M., Vogels, J., & Demberg, V. (2016). Salience and attention in surprisal-based accounts of language processing. *Frontiers in psychology*, *7*.

Zentall, T. R., Wasserman, E. A., Lazareva, O. F., Thompson, R. K., & Rattermann, M. J. (2008). Concept learning in animals. *Comparative Cognition & Behavior Reviews*.

Zhang, L., & Pylkkänen, L. (2015). The interplay of composition and concept specificity in the left anterior temporal lobe: An MEG study. *NeuroImage*, *111*, 228-240.

Zhang, S. P., Bandler, R., & Davis, P. J. (1995). Brain stem integration of vocalization: role of the nucleus retroambigualis. *Journal of Neurophysiology*, *74*(6), 2500-2512.