View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Scholarship@Western

Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

8-18-2014 12:00 AM

# Language, Mind, and Cognitive Science: Remarks on Theories of the Language-Cognition Relationships in Human Minds

Guillaume Beaulac
*The University of Western Ontario*

Supervisor
Christopher D. Viger
*The University of Western Ontario*

Graduate Program in Philosophy
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy
© Guillaume Beaulac 2014

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Philosophy of Mind Commons

**Recommended Citation**
Beaulac, Guillaume, "Language, Mind, and Cognitive Science: Remarks on Theories of the Language-Cognition Relationships in Human Minds" (2014). *Electronic Thesis and Dissertation Repository*. 2288.
https://ir.lib.uwo.ca/etd/2288

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

# LANGUAGE, MIND, AND COGNITIVE SCIENCE: REMARKS ON THEORIES OF THE LANGUAGE-COGNITION RELATIONSHIPS IN HUMAN MINDS

(Thesis format: Integrated Articles)

by
Guillaume Beaulac

Graduate Program in Philosophy

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Guillaume Beaulac 2014

# Abstract

My dissertation establishes the basis for a systematic outlook on the role language plays in human cognition. It is an investigation based on a cognitive conception of language, as opposed to communicative conceptions, viz. those that suppose that language plays no role in cognition (its only role being to externalize thought). I focus, in Chapter 2, on three paradigmatic theories adopting this perspective, each offering different views on how language contributes to or changes cognition.

In Chapter 3, I criticize current views held by dual-process theorists, and I develop a picture of the complex interaction between language and cognition that I deem more plausible by using resources from the literature on the evolution of the faculty of language. Rather than trying to find one general explanation for all cognitive processes, I take seriously the idea that our mind is composed of many subsystems, and that language can interact and modify each in different ways. There is no reason offered in the empirical literature—besides maybe parsimony—that suggest that language has to interact in the same ways with all cognitive processes. Yet, this is seemingly taken for granted, especially within dual-process approaches.

On my view, it is a central requirement for a theory of the role of language in cognition to explain how language might have effects, at once, on and within various parts of cognition. In Chapter 4, I explore how this framework can modify how we think about some experiments in psychology, specifically in research on categorization. My idea is that language, once it (or any possible primitive forms) evolved, changed how some cognitive capacities worked and interacted with each other, but did so in more than one or two ways. Cognitive systems are changed in very different ways—sometimes the transformation is very subtle, such as our way of forming categories by using how similar objects are, while other times it is deep and changes the very way the system works.

# Keywords

language, cognition, thought, dual-process theory, architecture of mind, cognitive architecture, concepts, philosophy of psychology, philosophy of mind, philosophy of science.

# Co-Authorship Statement

Chapter 4, “"Concept" Heterogeneity and Definitions”, has been co-authored with Pierre Poirier, Associate Professor and former Director of the Institut des Sciences Cognitives at Université du Québec à Montréal. The basis for this paper was first written as a commentary published in a book symposium on Edouard Machery’s (2009) *Doing without Concepts*. This paper was published in *Dialogue* (Poirier & Beaulac, 2011). Following Machery’s reply (2011), we continued the project and expanded greatly on this initial paper, hence the numerous references to Machery’s work and to his replies to some of the arguments. I have taken lead on the current paper: a first, shorter draft was submitted as a term paper for the course PHILOSOP 9605A (Concepts) taught by Angela Mendelovici in the Fall 2011 term, which we then revised. It was accepted for a full paper presentation at the 2012 meeting of the Society for Philosophy and Psychology.

As it is, my contribution is about 65% of the work on this version of the paper. Pierre Poirier wrote section 4.2, except 4.2.2 (which I wrote), but sections 4.6.1, 4.6.2, 4.7 and 4.8 were almost entirely (over 95%) written by me. The rest of the paper (4.1, 4.3, 4.4, 4.5, 4.6—except its subsections), including the overarching argument, including the planning of the sections and their role in the argument, is a collaborative effort.

Acknowledgments

Many would start by acknowledging the academic support they had during their PhD program but, in my case, this would be ignoring the person who suffered the most during the process leading to this dissertation. None of this would have been possible without the help, love and support that Evelyn McDuff has offered me since the beginning of this crazy adventure. Evelyn, thanks. I love you.

I want to warmly thank Chris Viger, my supervisor, for being open to the project and supportive of my twelve (plus) other projects these past four years. You allowed me to pursue academic freedom to its full extent, and I am grateful that you always were so confident I would get this done in time(!). I cannot stress sufficiently the impact of your ideas and your guidance on this project.

Gillian Barker and Robert J. Stainton, members of my prospectus committee, also played important roles in this process, and their comments during my defense of the project helped greatly shape the ideas in these pages. Thanks!

Thanks also to John Paul Minda who gave me detailed comments and suggested many references that deeply influenced the current version of Chapter 4.

Pierre Poirier, who taught me how to do (and love) research, is to be blamed for most of this (especially Chapter 4, which he co-authored!). Pierre, you convinced me that I had what it takes to embark on such enterprises and you offered me incredible support all the way through. Thanks for everything. Thanks for believing in me.

Thanks to Colin Allen, Normand Baillargeon, Frédéric Bouchard, Christine Daigle, Ronald de Sousa, Shannon Dea, Luc Faucher, Peter Godfrey-Smith, John D. Greenwood, Stevan Harnad, Tim Kenyon, Edouard Machery, Sara Magrin, Mathieu Marion, Genoveva Martí, Kate Norlock, Serge Robert, David Rosenthal, and Alain Voizard for being wonderful mentors and giving me much needed advice, and great opportunities, these past years. And thanks to Frédéric-I. Banville for being such a good friend and colleague. (Thanks, also, for having revised almost all of this ... and for commenting on just about every single thing I write.)

Thanks, finally, to my collaborators, people who offered comments on various parts of my work, who offered me support (emotionally and academically), who participated in reading groups (CRISCo and others), and who shared scotch and beer with me. So many more friends and colleagues should be named here but I promised myself to keep this short (I failed). I insist. I learned as much (if not more) from interacting with you, dear friends and colleagues, than I did collaborating with my advisor and other faculty members.

To anyone thinking about grad school: keep in mind that you are not only choosing a school, you are choosing colleagues who will play a huge role in your life as a grad student. And I'm happy that Western was a great choice in this regard.

Thanks also to the Rotman Institute of Philosophy for fostering an amazing research environment. My research also greatly benefited from research visits at the Department of Philosophy of the Graduate Center at City University of New York in the Spring 2013 term with Peter Godfrey-Smith and John D. Greenwood, at the Center for Philosophy of Science at the University of Pittsburgh in the Spring 2012 term with Edouard Machery and at the Université du Québec à Montréal's Canada Research Chair in Cognitive Science during the Summer 2011 term with Stevan Harnad.

This research was supported by the Social Sciences and Humanities Research Council (Joseph-Armand-Bombardier Doctoral Scholarship 767-2010-2249, and Michael-Smith Foreign Study Supplement 771-2012-0007), the Ontario Graduate Scholarships program, and the Schmeelk Canada Foundation.

Je m'en voudrais de ne pas conclure en offrant quelques mots de remerciement en français à mes parents, Linda Charron et Roch Beaulac, et à ma sœur, Amélie Beaulac, pour leur amour et leur support depuis ... le tout début de tout ça. Merci. Je vous aime.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1 : Introduction

*War is what happens when language fails.*

Margaret Atwood (1998)

# 1 Introduction

While descriptions of the role of language in human cognition are often restricted to its communicative role[1], many theories suggest that it contributes to cognition in various ways, and might even profoundly modify the way cognition works. This is more than to say that "language is an internal "instrument of thought"" (Berwick, Friederici, Chomsky, & Bolhuis, 2013, p. 91), however. It means that language can *indeed* be an instrument, although not necessarily only an internal one, but it can also be an instrument in a transformative sense, viz. its very use can change how cognition works. How language makes such contributions, and which capacities it interacts with—what it modifies, enhances or adds, however, are questions that need to be examined more closely. This is, at least, the claim I will be arguing for in these pages.

These are very broad issues, addressed by many approaches within philosophy (at various moments in its history) and elsewhere, which can hardly be handled within a single project. My intended goal is to contribute to the way these questions are framed within contemporary work in cognitive science. I propose, first, an examination of many theories of the role of language in cognition using the perspectives provided by three paradigmatic views in philosophy. I don't expect these three views to represent all possible positions; I prefer to see this as a first exploration of the landscape and some related, multifaceted, issues. This exploration in itself is still valuable as I expect most theories to fall somewhere

1. There *is* a communicative role for language, but whether this role is the primary role of language or a by-product that later came along for a ride (a spandrel) is not a question I will address here. See Pinker & Jackendoff (2005) for a discussion of this point.

within this landscape, and any theory that would *not* fit in this landscape would be interesting in *how it differs* from these paradigmatic views.

The views I will be examining are Carruthers' (2006, 2013a, 2013b) view of language as the way modules interface in a global workspace, a version of Clark's (2008; see also Donald, 1991) scaffolding hypothesis where I also discuss contemporary variations on Whorf's (2012) seminal proposal, as defended by, among others, Colombetti (2009), and Boroditsky & Prinz (2008), and Dennett's (1991, 1994, 2009a) rewiring thesis.

More than merely defining a landscape, I will argue that the weakness of the aforementioned frameworks lies in their attempt to explain what *the* (single) role of language in cognition is. I will make a case for a view where language plays multiple roles, roles that are well described in one or more of these paradigmatic approaches. Looking at some of the phenomena under study using this perspective, I think, helps to reinterpret some experimental results and might make it possible to see new dynamics emerge when it comes to understanding how the mind's processes are organized.

From this standpoint, I examine (Chapter 3) the dual-process theory framework (Evans & Frankish, 2009; Evans & Stanovich, 2013; Evans, 2008; Gawronski & Creighton, 2013; Kahneman, 2011; Stanovich, 2011) and discuss some of its limitations on how language is understood therein. Rather than trying to find *one* general explanation or *one* general framework for all cognitive processes, something which the authors above spend much energy doing—as this is seen as one of the major perks of dual-process views, I take very seriously the idea that our mind is composed of many subsystems, and that language can interact and modify each one in different ways. There is no reason offered in the empirical literature—besides maybe parsimony—that would suggest that language has to interact in the same ways with memory systems, categorization processes, reasoning skills, and other cognitive processes. Yet, this is seemingly taken for granted. The dual-process perspective is representative of such views—the macro categories ("Type 1 and Type 2 processes", "System 1 and System 2") are themselves too vague. There are very good reasons to think that there is more than *one* part to each of the categories mentioned above; memory is well-known to have its own distinctions, and the same goes for many other fields of inquiry within psychology.

On my view, it is a central requirement for a theory of the role language in cognition to explain how language might have both effects, on and within various parts of cognition. Then, in Chapter 4, I enquire how this framework might modify how we think about some experiments in psychology. My idea is that language, once it (or any possible primitive forms) evolved, changed how some cognitive capacities worked and interacted with each other, but did so in more than one or two ways. Cognitive systems are changed in very different ways—sometimes the transformation is very subtle, such as our way of making categories by using how similar objects are, while at other times it is deep and changes the very way the system works, such as the trade-off that seemed to occur in memory, such that our visual memory is very poor, unlike that of, e.g., chimpanzees, but our long-term and episodic memory systems seem much more efficient and precise than that of other species. Chapter 4 explores this hypothesis from the point of view of the psychology of concepts, or *concept science*, as we refer to it in these pages. The reason motivating this expression is that I believe that, down the road, psychology will not have all the necessary resources to explore the diversity of phenomena at play under the general label “concept”.

The aim is not to settle what the role of language in cognition is—a task well beyond the scope of a dissertation—rather it is to provide much needed conceptual clarification. My approach suggests new possibilities for the relation between language and thought that can inform research in philosophy of psychology and cognitive psychology. Each of the chapters of this dissertation, written as a standalone paper, accomplishes part of this overarching goal.

## 1.1 A case motivating this research

A case presented by Donald (1991), based on the description in Lecours & Joanette (1980), can illustrate how ideas about the role language is supposed to play change how we look at some of the data[2]. In their paper, Lecours & Joanette describe the case of Brother John, a Francophone unilingual Québécois suffering from paroxysmal aphasia, “without,” as the authors state, “modification of consciousness” (Lecours & Joanette, 1980, p. 1), due to epileptic spells. Brother John’s aphasia is characterized by short and long spells during which he loses the capacity to use and understand language—the spells more or less (and a lot stands on this “more or less”) shut down language processing. Brother John is still able, during those spells, to carry out complex social interactions that seemingly require the use of symbolic capacities, usually deemed to be among those that language is necessary for. Moreover, he is aware of his disability and he is usually able to cope with the

2. Thanks to Peter Godfrey-Smith for pointing out this example.

situation, finding ways not to "inconvenience himself or others" (Donald, 1991, p. 84).

Brother John had been having these spells for some time when Lecours & Joanette started to work with him. He was used to the spells, was able to anticipate them, and he was sometimes able to arrange for one of his colleagues to cover for him when they happened. One striking example that Donald (1991) emphasizes happened during a trip to Switzerland. During a train ride, Brother John realized one of his long spells, one that can last up to 8 hours, was about to start. When he got out the train, in a town he never visited before and despite being unable to use language in any form, he was still able to get to a hotel, using mimes to get a room. He ordered food by pointing at something at random (hoping he would like it) in what he believed would be the hors-d'oeuvre section of the menu. When the spell was over he remembered what happened during the spell in every detail.

For Donald, this case shows that "despite the complete absence of language, internal or external" (Donald, 1991, p. 85), Brother John is able to act in ways that seem clearly to be uniquely human: complex planning, coherent thought, memory formation, ability to interact socially and to recognize faces, places and even some things very specific and contextual such as sections on an unknown restaurant's menu. His episodic memory also remains intact, like some gestural abilities and practical knowledge. Hence, despite the absence of language, Donald thinks Brother John is able to exhibit abilities that go far beyond anything other primates can do. Language, of course, can help cognition in many ways and has a strong role in social transmission and in communication, but he believes that "[l]anguage was obviously not the vehicle by which he assessed events, formulated plans, and evaluated his own responses" (Donald, 1991, p. 85).

Dennett (1994), however, is not impressed by this example. Because Brother John *normally* has language, Dennett argues that his cognition benefits from the shaping of language *even during his spells*. In other words, it is not true, according to Dennett, that we can talk about a *complete* absence of language in this case. Brother John's cognition, even when, e.g., inner speech is not available, is deeply transformed by his use of language in normal circumstances. This, as I will discuss in Chapter 2, is a marked difference between the rewiring and the scaffolding theses. Moreover, since Brother John can anticipate his spells, it is possible that he might do a lot of planning beforehand. Dennett insists:

> These varieties of language-less thought, like barefoot waterskiing, may be possible only for brief periods, and only after a preparatory period that includes the very feature whose absence is later so striking. (Dennett, 1994, sec. 5)

Yet, it is not a decisive narrative which proves that language is transformative of every cognitive ability in the way Dennett thinks it is.

Brother John's story, however, might be harder to account for in Carruthers' (2006, 2013a) framework since the rehearsal of inner speech actions plays a very large role in how he explains the unique flexibility of human behavior. For Carruthers, who does not (so far) discuss this specific case, explaining Brother John's case requires a modification to his view, where language could have lasting effects. Yet, some aspects of the case do speak in favor of Carruthers' view: Lecours and Joanette (1980) also mention that Brother John uses some labels, even during his spells—something Dennett (1994) notes to strengthen his own position—such as "turn", "push", etc.

I believe that these theories, briefly introduced above, all point towards interesting aspects of the phenomena but that, in the end, the pictures they present of the mind have trouble accounting for some *other* aspects of the same phenomena. The task of this dissertation will be to shed some light on the issue and, hopefully, find a way to better understand what is going in such cases.

## 1.2 The philosophical project

As this is a project in philosophy of psychology and, more generally, cognitive science, I will proceed to analyze theories widely influential in cognitive science concerning the relation between language and thought in terms of the concepts they use, their methodological approaches, and the consequences this might have on how we understand the human mind. I do not believe it is necessary to *justify* this project as being philosophical in nature (cf. Dennett, 2009b), but I would like to point out a few general implications I see down the road for this kind of project. First, I would like to draw attention to an example in the history of philosophy that illustrates the importance of such questions in the tradition by looking at Condillac's views on the matter. Although my main research question, viz. how language modifies cognition, is rarely formulated in this exact way, I think there are precursors in the history of philosophy; my example being from the modern period. Then, I will mention a few questions—some from the evolution of language literature and others from work in animal cognition—about how this is one but many examples of how philosophy can be important within cognitive science.

### 1.2.1 Language and cognition in Condillac

In his attempt to understand the human mind, and to define human nature, Étienne Bonnot de Condillac discusses evidence from observations of nonhuman animal behavior. He recognizes the striking similarities in how humans and animals act in their respective environments. He also accepts that analogies between human and nonhuman behavior are useful and informative, and he uses them abundantly in his work. He thinks the study of animals can be enlightening and help us to understand human cognition because animal behavior can be informative of our own—mostly because Condillac thinks both types of cognition work according to some of the same principles, revealed in simpler forms in animals.

In his approach to the study of the animal mind, Condillac adopts Hume's principle of similarity enunciated in the *Treatise* (Hume, 2011, para. 2.1.12.2), that is "where the structures of parts in brutes is the same as in men, and the operation of these parts also the same, the causes of that operation cannot be different". In the *Traité des animaux*, Condillac (2004) explains that the idea of the animal-machine is absurd for him: central to his views of both humans' and animals' mental life is the idea that sensations are—to put it into contemporary words—functionally relevant in behavior. Without sensations, behavior would not be the way it is.

From this foundation, and by stating the importance of sensations in his grasp of both humans' and animals' lives, Condillac ties the views he develops in the *Traité des animaux* to those of his *Traité des sensations* (Condillac, 1984), a treatise where he offers a reconstruction of our understanding of human cognition building on the principle that humans tend to maximize pleasure and minimize pain, and that the only built-in processes are basic perceptions from the sense organs. That is, he rejects that there are any innate abilities beyond those. Remembering, reasoning, abstracting, even focusing attention are only learned, they are taught by experience alone (Falkenstein, 2010). Development of a higher cognitive faculty starts from the ability of focusing attention on various objects perceived. Sensations are also richer than what is perceived: cognition then has a lot to do with the discovery of what is "already present in [...] sensations" (Falkenstein, 2010, sec.

7.2). By a complex process of learning and habituations (what Condillac calls the acquisition of "habitudes" (habits)), animals and humans come to, e.g., anticipate some events to follow their current sensations and act accordingly. This brings Laupies to state that what comes naturally is "the habit of which we forgot the origin" (Laupies, 2004, p. 116, my translation). It is important to emphasize the following point: there is not any difference, *of kind or of degree*, between sensations humans and animals have. Condillac goes even one step further. For him, the soul cannot be distinguished from the body, the distinction being "neither conceivable nor useful" (Laupies, 2004, p. 106, my translation).

Despite these numerous similarities, there is certainly the appearance of much dissimilarity between humans and nonhuman animals, and Condillac has to account for this as well. He invokes two general principles: the first one is tied to the different needs of different species—intrinsically related to the cognitive capacities *developed*—and the second one is explained by Condillac's views on language and communication. I will not, here, go into the details of the first reason, but the second one is most interesting for my current purposes.

This second one has to do with language, a central feature of Condillac's views on humans. As Garrett explains, "the comparison of human and animal sensitive, imaginative, and intellectual capacities resulted in a language-centered account of human nature" (Garrett, 2006, p. 164) and this is the very element that sets human beings apart from animals—this is where the difference in *kind* rises for Condillac. It adds "an intersubjective social and historical medium" (Garrett, 2006, p. 164) to the psychological theories proposed by earlier authors. This is a central and very rich idea. This is not without tensions as, again, in principle, there is nothing that would make it impossible for animals to achieve this kind of cognition if they had *the relevant needs* as they also have some forms of language in Condillac's view. In practice, however, it does place humans into a different realm, a cultural and social one—and Condillac brings forward the idea that humans are superior to animals. Language, according to what Condillac says in his *Logique* (Condillac, 1780), is what creates rational thought. The idea is *not* that no form of reflection is available to animals, but language brings it to another level; thought in animals lacks precision and consists mostly of instincts. Humans' reflection goes beyond that of animals mostly because it has this combination of instinct and reason, closely interlinked; reflection in the human case is "solicited above and beyond habits by the diversity of the needs, it can be abstraction" (Laupies, 2004, p. 124, my translation) which also means that humans can think beyond what is directly in their surroundings.

According to Coskies, in Condillac's framework, "[i]f animals have less language and less reason, it is only because their physical condition has not compelled them to attain more" (Coskies, 2003, p. 69). Dagognet describes the difference between humans and animals by the fact that "the animal is born with less organic means, that are less differentiated, and—mostly—that are adapted to the animal's functions of autopreservation" (Dagognet, 2004, p. 125, my translation). He adds that, in Condillac's *Traité des animaux*, the complexity of human beings sensory apparatus makes humans relatively less dexterous and more vulnerable early in life. However, it allows for and encourages, ultimately, the human greater need for more habits and—above all—encourages curiosity, "this insatiable need for knowledge" (Condillac, 2004, para. 2.5.14, my translation). These are the conditions needed for the development of human language.

Once human language arises, it allows for a whole different kind of cognition, a "higher type of understanding" (Coskies, 2003, p. 66), theoretical knowledge. Condillac even goes as far as identifying the ability of using language properly with the capacity to reason correctly in his *Logique*; clear connections are made by putting linguistic expressions together. Human language also permits information sharing and the development of culture, but also the differentiation of individuals within a society. One of the limitations of animals is that each individual has to learn everything by itself: "[the beasts] are more or less limited to the pieces of knowledge that each individual can acquire by itself" (Condillac, 2004, para. 2.4.17, my translation). In human societies, knowledge can be shared—the examples of such social learning in nonhuman species being impressive but still limited. Above all, theoretical knowledge allows for the knowledge of God and for morality to emerge, the clearest signs of the superiority of humans in the animal realm. Condillac's view is well summarized by Garrett:

> Each animal, however social, begins anew and responds to experience no faster than previous generations, unable to learn from the species's mistakes and successes through a common storehouse of language. Man, on the other hand, accumulates reflective knowledge, which, although derived from the senses, is able through language to expand and progress into an open-ended future. (Garrett, 2006, p. 164)

The difference in kind, therefore, emerges with the appearance of the capacity for language. Its rise is explained by a difference in degree, but—once it is present—creates the wide gap between humans and animals Condillac was looking for. What exactly makes humans unique despite the similarities and the nature of the differences (differences in degrees) that they observe? The answer, for Condillac, lies in language and how it changes some aspects of cognition.

This historical take on the kind of problem I am interested in shows two things: first, the question has not changed much. What makes humans unique? If it is language that makes humans unique in the animal realm, how does language contribute to human uniqueness? The second thing it reveals is that there are insights for contemporary problems in cognitive science to be found in the history of philosophy. While this is not exactly a new observation, the way Condillac's views connect with contemporary problems is striking. As we will now see, Condillac was defending a kind of view that is nowadays defended by the likes of Toates (2006) and Fitch (2010), and he was doing so in light of his observations of nonhuman species behavior. Although I will not be coming back to these historical questions here, I think it will be useful to keep in mind how Condillac was understanding and explaining these issues more than 200 years ago.

### 1.2.2 Language, cognition, and research on the origins of language

In his recent overview of the research on the origins of language, Fitch holds that there is no doubt that nonhuman animals have sophisticated cognition, viz. sophisticated cognition without language (Fitch, 2010, pp. 171–172). In fact, Fitch is so optimistic about recent discoveries in animal cognition that he suggests a new take on the problem of the evolution of language.

Hauser, Chomsky and Fitch (2002) suggest an interesting distinction between the Faculty of Language in the Broad (FLB) and in the Narrow (FLN) sense. FLN refers to the part(s) of the language faculty that is (are) uniquely human and FLB refers to the capacities necessary for language that are shared with nonhuman animals. FLB includes various elements like the vocal chords and the lungs, but also many cognitive capacities necessary for language, including memory, categorization system(s), etc. More recently, Fitch (2010) suggests an interesting take on this reflection: as first hinted at in Fitch, Hauser, & Chomsky (2005), FLN might be an empty set, that is, nothing would be part in and by itself, of FLN, not even recursion (which was the previous candidate for FLN). Yet, if FLN is empty we still have to explain the—what appears to be unique in the animal realm—case of language.

Fitch's suggestion is that what is unique about human beings is not necessarily any cognitive capacity but rather the particular combination of specific capacities. If Fitch is right, this would mean that every part that is needed to have a human-like language faculty could be found in a number of nonhuman animals in different combinations, viz. no capacity would have to be unique to human beings, but these parts are never found in the peculiar arrangement that happens to work in the human case. This hypothesis calls for an investigation of the capacities necessary for language and how they are found in nonhuman animals.

This is where my own research program comes in, as I add a twist to Fitch's suggestion. Here, the difficulty is that there is no framework to think about what changed once language started to interact with, or to modify, cognition (or parts of it). What I mean by this is that the ideas suggested by Fitch and his colleagues do not provide us with tools to go in the details of the role language plays once it appears in cognition. Put differently: how does language affect each part of FLB and FLN once it sets in? Are these parts all affected in the same way?

My motivation is that there is an apparent incompatibility between Fitch's proposal and dual-process theories, at least in their standard incarnations (as I will detail in Chapter 3). What is needed here is an account of the architecture of mind that can make sense of Fitch's idea, and I believe that a weakened version of dual-process theories would do the trick. A proposition like Fitch's would work *with* the dual-process view only, here, some parts of FLB are akin to Type 1 processes while other parts are closer to the characterization of Type 2 processes—or their closest nonhuman version.

Let me illustrate this idea: even though Fitch holds the view that there is no uniquely human process, he lacks an adequate explanation of how these mechanisms all come together and what, once a species acquires language, changes. For instance, it is not clear what a component of FLN (if FLN ends up not being an empty class) would be like, and what effect it would have on cognition more generally. If recursion happens to be the "missing piece" as suggested by Hauser and his colleagues (2002), how does it affect or change the various components of FLB? Would it be possible that, once FLN sets in, it radically changes the way cognition works (in the way Dennett (1991) envisioned it)? This shortcoming opens the door to more skeptical views of the cognition of nonhuman animals, such as Penn, Holyoak & Povinelli's (2008).

According to Penn and his colleagues (2008), there is a strong discontinuity between humans and nonhuman animals. After listing many of the observed differences, they add that, while there is a "profound biological continuity between human and nonhuman animals", in fact, this apparent continuity "masks an equally profound functional discontinuity between the human and nonhuman mind" (Penn et al., 2008, p. 110). Toates (2006) makes the opposite point as he thinks that we can see most of these cognitive capacities as "uniquely developed" in humans—but only a few capacities are unique to human beings. This is especially interesting as there seems to be an agreement on the data, but the interpretation varies wildly.

Yet, what appears to create the "profound functional discontinuity" described by Penn and his colleagues, might not be so profound in the end. My worry is that it is hard to compare and evaluate these interpretations of the data in a context where we do not know what role language plays. What if the presence of language, alone, explains all of these differences? Much about these unique abilities seems to be, at the very least, dependent on the capacity to deal with abstraction and logical relations—which are strongly tied to capacities that seem to depend on language.

Without having an idea of which parts of cognition language modifies, I hardly see how the discontinuity can be assessed. It might be true that language radically modifies the type of cognition observed, but the change could be smaller than what Penn and his colleagues seem to suppose. Yet, Fitch might be wrong in thinking that there is nothing crucial that differentiates humans from other animals; there might be something missing from the FLB mechanisms, something—yet to be identified—that would create a discontinuity. There might also be degrees of interaction, and language might interact in different ways with different cognitive subsystems.

This perspective helps identify problems and difficulties encountered in the animal cognition and in the architecture of mind literatures. It might even bring about some suggestions for changes in methodology, and more precision in the concepts used to talk about these phenomena, especially when cognitive capacities (partly) shared between humans and other animals are under scrutiny. This is what this research project is all about. Even though I might not be able to fully provide answers, it is my belief that the tools I lay out in the following pages are at least helpful in advancing the discussion, and this is how the three following chapters fit together. Their goal is to advance various aspect of these discussions. If I am right, each of these chapters would contribute to make steps in the direction of solving the very hard questions mentioned so far—or maybe some simpler, easier, ones that will help in answering them.

## References

Atwood, M. (1998). *The robber bride*. New York: Anchor Books/Doubleday.

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. doi:10.1016/j.tics.2012.12.002

Boroditsky, L., & Prinz, J. J. (2008). What thoughts are made of. In *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches* (pp. 98–115). New York: Cambridge University Press.

Carruthers, P. (2006). *The architecture of the mind : massive modularity and the flexibility of thought*. New York: Oxford University Press.

Carruthers, P. (2013a). Animal minds are real, (distinctively) human minds are not. *American Philosophical Quarterly*, *50*, 233–247.

Carruthers, P. (2013b). The fragmentation of reasoning. In P. Quintanilla (Ed.), *La coevolución de mente y lenguaje: Ontogénesis y filogénesis*. Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú. Retrieved from http://faculty.philosophy.umd.edu/pcarruthers/The%20Fragmentation%20of%20Reasoning.pdf

Clark, A. (2008). *Supersizing the mind : embodiment, action, and cognitive extension*. New York: Oxford University Press.

Colombetti, G. (2009). What Language Does to Feelings. *Journal of Consciousness Studies, 16*(9), 4–26.

Condillac, E. B. de. (1780). *La Logique, ou les premiers développemens de l'art de penser*. Paris: L'Esprit & Debure l'aîné. Retrieved from http://visualiseur.bnf.fr/CadresFenetre?O=NUMM-80138&M=tdm

Condillac, E. B. de. (1984). *Traité des sensations. Augmenté de l'extrait raisonné*. Paris: Fayard.

Condillac, E. B. de. (2004). *Traité des animaux*. (M. Malherbe, Ed.). Paris: J. Vrin.

Coskies, R. C. (2003). Condillac: Language, Thought, and Morality in the Man and Animal Debate. *French Forum, 28*(1), 57–75.

Dagognet, F. (2004). *L'animal selon Condillac: une introduction au "Traité des animaux" de Condillac*. Paris: Vrin.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

Dennett, D. C. (1994). Language and intelligence. In J. Khalfa (Ed.), *What is Intelligence?* Cambridge: Cambridge University Press. Retrieved from http://ase.tufts.edu/cogstud/papers/rolelang.htm

Dennett, D. C. (2009a). The Cultural Evolution of Words and Other Thinking Tools. *Cold Spring Harbor Symposia on Quantitative Biology*. doi:10.1101/sqb.2009.74.008

Dennett, D. C. (2009b). The Part of Cognitive Science That Is Philosophy. *Topics in Cognitive Science, 1*(2), 231–236. doi:10.1111/j.1756-8765.2009.01015.x

Donald, M. (1991). *Origins of the modern mind : three stages in the evolution of culture and cognition*. Cambridge (MA): Harvard University Press.

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology, 59*(1), 255–278. doi:10.1146/annurev.psych.59.103006.093629

Evans, J. S. B. T., & Frankish, K. (Eds.). (2009). *In Two minds : Dual processes and beyond*. New York: Oxford University Press.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science, 8*(3), 223–241. doi:10.1177/1745691612460685

Falkenstein, L. (2010). Étienne Bonnot de Condillac. *Stanford Encyclopedia of Philosophy* (Fall 2010 Edition). Retrieved from http://plato.stanford.edu/archives/ fall2010/entries/condillac/

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

Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition, 97*(2), 179–210. doi:10.1016/j.cognition.2005.02.005

Garrett, D. (2006). Human Nature. In K. Haakonssen (Ed.), *The Cambridge History of Eighteenth-Century Philosophy* (Vol. 1, pp. 160–233). New York: Cambridge University Press.

Gawronski, B., & Creighton, L. A. (2013). Dual-Process Theories. In *The Oxford handbook of social cognition*. New York: Oxford University Press. Retrieved from http://publish.uwo.ca/~bgawrons/documents/GC_Oxford.pdf

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. doi:10.1126/science.298.5598.1569

Hume, D. (2011). *A treatise of human nature : a critical edition*. (D. F. Norton & M. J. Norton, Eds.). Oxford: Clarendon Press.

Kahneman, D. (2011). *Thinking, fast and slow*. Toronto: Doubleday Canada.

Laupies, F. (2004). *Traité des animaux d'Étienne Bonnot de Condillac: leçon philosophique*. Paris: Presses universitaires de France.

Lecours, A.-R., & Joanette, Y. (1980). Linguistic and other psychological aspects of paroxysmal aphasia. *Brain and Language, 10*(1), 1–23. doi:10.1016/0093-934X(80)90034-6

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31*(2), 109–130; discussion 130–178. doi:10.1017/S0140525X08003543

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition, 95*(2), 201–236. doi:10.1016/j.cognition.2004.08.004

Stanovich, K. E. (2011). *Rationality and the reflective mind*. New York: Oxford University Press.

Toates, F. (2006). A model of the hierarchy of behaviour, cognition, and consciousness. *Consciousness and Cognition, 15*(1), 75–118. doi:10.1016/j.concog.2005.04.008

Whorf, B. L. (2012). *Language, thought, and reality: selected writings of Benjamin Lee Whorf* (2nd ed.). Cambridge (MA): MIT Press.

# Chapter 2 : First paper

*Le problème du langage se situe entre*
*la philosophie et la psychologie.*

Maurice Merleau-Ponty (2001)

# 2 The Role of Language and the Architecture of Cognition

## 2.1 Introduction

In their 1999 introduction to philosophy of psychology, George Botterill & Peter Carruthers discuss the place of natural language in thought (Botterill & Carruthers, 1999, pp. 208–225) or, as I will put it, the role language plays in cognition[3]—how language can transform aspects of cognition. They identify four *cognitive* conceptions of language, i.e. conceptions of language that assume that language's role is not limited to communication. I will detail these roles shortly. These conceptions are opposed to the classic Humboldtian view that language and thought are identical. They also go beyond conceptions about the kind of role language can have that were made popular by scholars such as Noam Chomsky. For example, Berwick, Friederici, Chomsky & Bolhuis write that "language is an internal "instrument of thought"" (Berwick et al., 2013, p. 91), but it is implied here that language does not have a transformative role, beyond—in some cases—the construction of a thought. It is theoretically possible in Berwick and colleagues' model that the externalization of natural language modifies in some ways syntax

3. I will be using "thought" and "cognition" interchangeably, but I have a strong preference for "cognition". The first reason is that the philosophical literature on this question uses the word "thought" when it comes to discuss the interaction between language and cognition, and that "cognition" is generally used in cognitive science to talk about what philosophers mean when they use "thought". The second reason is that "thought" is usually associated with inner speech—Carruthers' work, as we will see, is a good example of this, while cognition is generally understood to have a wider extension. With "thought", "cognition" and "cognitive processes", I refer to processes that manipulate, store and transform information—whether it is done consciously or not, and in (human) minds or not. Note that artificial systems can be "cognitive" in this sense—they can have thoughts, and I am willing to bite this bullet in the present context.

but not much is said, so far, in this type of linguistics framework about this possibility. A similar idea about the role of language has been put forward by de Sousa in a similar fashion: “As we shall see, language actually enlarges the domain of possible thoughts.”[4] (de Sousa, 2007, p. 72)

This is a controversial thesis. Whether the language faculty and some of its apparatus has influence on thoughts, or whether the structure of thoughts influence aspects of the language faculty is not an easy problem to solve. Now famously, a good part of the debate between Everett (2005, 2012) and Chomsky (2013; Hauser et al., 2002) has been whether or not recursion—the ability in a language to embed part of a sentence or of a thought in another one—is a universal characteristic of human languages and whether or not recursion as a cognitive ability is unique to the human species. There is no question that recursion is a common human characteristic. Recursive thinking is something we commonly do, and the Pirahã people that Everett studied are no exception. The syntax of their language, however, arguably does not have this property, although it is used in their stories. Everett’s question, then, is whether or not “the linguistic engine room [is] the same for all languages” (Everett, 2012, p. 281). Recursion, as a cognitive capacity, could originate in cognition, and the recursion in natural languages would be—in a sense—parasitic on it, or it could originate in grammar, or processes unique to the

---

4. This illustrates well how closely “cognition” and “thought” are related in the philosophical literature: just before embarking on a discussion of the modularity of mind, and after discussing how language changes intentionality (more on this below), de Sousa writes that he will be exploring “the relation of language to some of the other essential capacities of thought” (de Sousa, 2007, p. 76). Chief among these other capacities is “[mitigating] the isolation of our mental modules”.

Another example is that, under the heading “The place of natural language in thought”, Botterill & Carruthers’ (1999) first few words are “When the question of the place of natural language in cognition has been debated by philosophers [...]” (Botterill & Carruthers, 1999, p. 208).

human species and its faculty of language, and other cognitive processes would be parasitic on this. If this latter option ended up being right, grammatical recursion could modify how cognition works, and the lack of this cognitive capacity would likely change how these people's cognition works, since they are lacking linguistic recursion. If the former option is right, we would likely find other examples, or at least precursors, of recursion in nonhuman animal species. Although I will not go into the details of this debate, I think it illustrates well what kind of issues depend on having a clear account of the role of language in cognition—or at least good guidelines!—to think about these questions. However, a related issue in this debate between Everett and Chomsky, one that I go into in 2.2.1 and in 2.3.1, is to investigate what kind of cognitive tool language might be.

In this paper, I first detail Botterill & Carruthers (1999) four cognitive conceptions of language. I then develop (2.3) a different way of dividing the landscape of possible positions. I argue that one of the distinctions—language as the vehicle for conscious thought (second) and language for explicit conscious thought (third)—should be grouped as one conception (represented, among others, by Carruthers' (2006, 2013a, 2013b) view). I also include in the first view what I call contemporary variations on Whorf's (2012) controversial thesis that language modifies how we think (an option considered by Carruthers (2012)). More generally, I argue that the way these views are presented makes them incompatible although the authors sometimes do not realize it. I offer an analysis of this incompatibility and then offer a framework to resolve these tensions. Moreover, I suggest that Carruthers' (2012) idea that these different views are different points between two extremes—between the idea that thought is fully language independent and the idea that it is fully language dependent—should be replaced by a pluralistic approach. In 2.4, I will suggest a framework in which we can make compatible the different roles of language reviewed in 2.3.

## 2.2 Four cognitive conceptions of language

Botterill & Carruthers (1999) present four options when it comes to understanding the role of language in cognition. In this section, I detail these four options before criticizing the way the distinction is drawn and show how I modify this initial sketching of the landscape. I must first note that, for an unknown reason, this initial, four-fold division excludes an influential perspective from the literature. Although it is sometimes ridiculed, the Sapir-Whorf hypothesis has been the grounds for many debates in the past century or so and, while most agree that it might be too extreme, it brings some interesting elements to the forefront. Seuren defines the Whorf hypothesis (WH) as the hypothesis according to which

> language influences [weak WH] or determines [strong WH] thought and because any natural language is a shared property of a speech community, language influences or determines culturally bound ways of thinking with all the consequences thereof. (Seuren, 2013, p. 29)

I will come back to the Whorf hypothesis in 2.3.1 but it is in an important sense a variation on Vygotsky's (2012) view. For now, I will focus on the four options presented by Botterill & Carruthers, ordered "from the weakest to the strongest" (Botterill & Carruthers, 1999, p. 211): the scaffolding view, the consciousness view, the explicit view, and the transformation view.

### 2.2.1 The scaffolding view

According to Botterill & Carruthers, "almost everyone" agrees with this view "to a greater or lesser degree" (Botterill & Carruthers, 1999, p. 211), suggesting that one can be on board with this view and still advocate for one of the others. The idea of the scaffolding view—inspired by Vygotsky (2012)—is that words and symbols can be used to help thought in various ways. Examples often given include numbering systems to help process arithmetic, especially complex operations, the process of writing down ideas to help memory[5], the process of repeating, over and over again, a sentence to remember it more easily, being told about a rule (e.g., instructions) or formulating one explicitly. On its face, this view is not very controversial and seems to be obvious, but it can be adopted more or less radically.

The less radical perspective on the scaffolding view is that it makes some operations easier or might help in realizing more complex versions of a task, or that it can make it possible to gain access to different information[6]. Some concepts might even be completely inaccessible without language—thinking about "July 13th, 1988" or scientific concepts developed by the community of scientists are examples of this. As Botterill & Carruthers explain, however, this only means "that language is *required for* certain kinds of thoughts" but not that it changes how information is processed, that it is "involved in" or that it serves as a "*representational* vehicle" (Botterill & Carruthers, 1999, p. 212). In other words, it does not mean that the

5. One can hold this view without adopting the more radical claims from the extended mind approach.

6. Even Fodor (1975) thought that there could be such an influence of language on cognition. In *The Language of Thought*, Fodor suggests that with words such as the logical terms introduced with natural language, we have access more easily to some expressions in mentalese:

> In the paradigm case—the use of terms in a natural language—this correspondence [between the use of certain words and a state of affairs] holds because the speaker knows and adheres to conventions that govern the language. For, as we shall see in Chapter 3, such conventions fundamentally *are* the rules which pair propositional attitudes like beliefs with the forms of words that express those attitudes. (Fodor, 1975, pp. 72–73)

Making this idea clear is the project Viger (2005) is pursuing, notably by suggesting that for some expressions in mentalese that are computationally harder to track, the natural language term might be necessary to entertain some thoughts. I will say more about in 2.2.5.

way a given process works is changed by language. From a more radical perspective, the view is that language makes possible new forms of thinking, i.e. that some thoughts would not be possible without specific linguistic tools. Under this more radical perspective, language makes it possible to build up thinking, as well as increase its expressive power, viz. for this more radical view, more thoughts can be thought once language is acquired.

Clark (2008) and Everett (Everett, 2005, 2012; Frank, Everett, Fedorenko, & Gibson, 2008) adopt a view that Botterill & Carruthers (1999) claim is in between these two extremes. They see language as a *cognitive tool*. Clark insists on our use of outside resources to enhance the way some processes work, whether through speech (saying something to remember it better), writing down ideas or building a variety of cognitive tools—from arithmetic to the scientific method. Language can "augment cognitive powers" (Botterill & Carruthers, 1999, p. 214). Everett highlights this point by arguing that language is a tool that allows memory to perform in better ways; in the experiment he conducted with Frank and colleagues (2008), he shows that the Pirahã speakers can compare and match exactly large sets equivalent in number, but that they have difficulties doing this task from memory. In this experiment, when the participants compare sets and are asked if they match, they are able to give the right answer if the sets in question are seen. When they try to do this task from memory, they fail. Everett explains this by the fact that the Pirahã do not have words for exact numbers—they can express small and large quantities, but they do not have words to count (such as numerals). They have basic numeracy abilities as well as the capacity to compare the sizes of sets, but this capacity fails them when they try to remember how big the sets were. Everett argues, from this experiment, that language has "a fundamentally compressive role" (Frank et al., 2008, p. 823) that helps remembering some information in the long-run, although it does not change how said information is first perceived (a change in perception would indicate a strong Whorfian approach, cf. 2.3.1). We might want to be careful with this characterization, however, since this view of language as a tool can be adopted from a different perspective. Chomsky writes that

> [l]anguages are not tools that humans design, but biological objects, like the visual or immune or digestive systems. Such organs are sometimes said to have functions, to be for some purpose. (Chomsky, 2013, p. 655)

This means that, for Chomsky, language is to be studied like an organ used as a tool and not engineered for specific needs, and not as an adaptable tool transformed by social interactions and situations. According to proponents of this latter view, adopted by Everett (2012) and many others, including Tomasello (2014)—a well-known anti-Chomskyan (Tomasello, 2009), that language is a cognitive tool means that language is constructed, like other tools are, by human communities to fulfill a variety of needs. I will come back to my own understanding of language as a cognitive tool in 2.3.1.

Finally, the latter, more radical option for the scaffolding view, is inspired from Vygotsky's view, where language is presented as an integral part of thinking and its development. By learning language, and receiving instructions, rules and other varied indications, Vygotsky claims that a child's development will be different. This means, for example, that a child who is given specific instructions would be able to learn things that would be inaccessible otherwise and do things which other children cannot do. Some versions of this view can be constrained to certain parts of the mind. Under some perspective, the executive function is seen as entirely dependent upon language being present—a view close to Carruthers' own (more on this in 2.3.2).

The degree of involvement for language in the scaffolding process varies. In the Frank and colleagues' experiment, is memory merely enhanced by the presence of language or is it constituted by language and the type of compression it makes possible? Allen (2014), pursuing the metaphor, writes that language is sometimes used, just like a scaffold is, to build cognitive structure and that it can happen, once the structure is up, that the scaffold need not be there to perform some tasks. His example is rote learning times tables in basic arithmetic. These are learned by repetitions and, for some individuals, these tables become internalized. Once they are, it is like a cognitive shortcut is installed in the mind of these individuals making it just as easy to answer "three times four" as it is to answer "eight times nine". The well-drilled individual, Allen writes, becomes very good at answering complex arithmetical questions without needing use of language to give an answer; the times tables "are scaffolding in the primary sense of temporary processes or materials that facilitate the construction of other structures" (Allen, 2014, p. 234). I will use this example and other similar ones to address the difference between two roles for language, one according to which it is used to enhance some thinking processes and another one according to which language can *transform* these processes in 2.2.4, and then further in 2.3.1 and 2.3.3.

### 2.2.2 Language as the vehicle for conscious thought

This idea, defended by Carruthers, is that "imaged natural language sentences, in 'inner speech' are the primary vehicles for *conscious* propositional thoughts" (Botterill & Carruthers, 1999, p. 217, my emphasis). This view, compatible with

Fodor's (1975) account of mentalese / language of thought, is that most of cognition happens via other means than natural language[7]. However, what is *conscious* happens via tokened natural language sentences. According to this position, this is the main contribution of language in evolution—inner speech plays a causal role in making thought processes, carried out in mentalese, consciously accessible. The view does not fall into epiphenomenalism, however, as this inner speech can enhance some thought processes that would be too hard to carry out in mentalese only. We saw examples of this in 2.2.1. As an example, repeating a telephone number mentally—for those who do not have mobile phones!—can help remembering it.

In recent work, Carruthers uses such a model to explain the difference between the two systems posited by dual-process theories (Evans & Stanovich, 2013; Evans, 2008; Gawronski & Creighton, 2013), System 1, an automatic, fast, and unconscious system having multiple parallel processes, and System 2, a controlled, slow, and conscious serial system. For Carruthers (2009), this division of the mind into two systems comes with multiple problems (see also Keren & Schul, 2009; Samuels, 2009), one of which is to explain *why* there would be two systems. The traditional perspective is the "default interventionist" where System 1 processes handle information by default, mainly because they are much faster and require less resources than System 2 processes, and System 2 processes will sometimes—under appropriate conditions—intervene to replace an output that came from the System 1 processes. The default interventionist approach thus posits that processes from both Systems can be active at once, but they cannot both command behavior

7. I use "natural language" here to contrast with "language of thought".

(so they will not be used at the same time). Carruthers rightly sees this as a problem. Why would, from the point of view of evolutionary constraints, two systems evolve "*alongside* each other, competing for control of the person's behavior" (Carruthers, 2012, p. 395). Having a whole new system evolving to monitor the first one seems a costly solution and Carruthers' proposal has the benefit of being more economical.

For him, then, the processes of System 2 are really "*realized* in those of System 1" (Carruthers, 2012, p. 395), viz. System 2 does not exist in and by itself. It is a result of the activation of numerous System 1 processes. The main advantage of this view is that the two systems are intertwined, and it becomes easier to explain how one can replace the other when it comes to process some pieces of information. The main cognitive tool driving these multiple iterations of System 1 processes is language, through the rehearsal of inner speech actions. By using inner speech, information outside of a module's domain is made available to it. This imaged rehearsal of action happening through System 1 processes is broadcast in a global workspace *à la* Baars (1997) where it then becomes available to other System 1 processes. Hence, the properties of System 2—including consciousness—are the result of these iterations of System 1 processes done mainly through language. The processing of System 1, in contrast, is done at a subpersonal level and, maybe, through mentalese.

### 2.2.3 Language as the vehicle for explicit thought

On the previous view, mentalese *can* be involved in thinking but does not *have* to be. The real vehicle of thought might be mentalese and, in such a case, natural language does not have a direct effect on cognition (hence the mention of epiphenomenalism above). On this third view, however, both conscious and non-conscious thought processes are carried out in natural language (and not in mentalese), through logical form (Chomsky, 1995). This logical form, not related to the kind of logical form we usually encounter in philosophy, is provided by the resources of the language faculty and might vary depending on the language that is learned. This variant of the view is thus stronger than the previous one, where unconscious processes are not carried out in mentalese. Logical form, for Chomsky, is "the level of linguistic representation [...] where the language faculty interfaces with central cognitive systems" (Botterill & Carruthers, 1999, p. 223), and thinking simply is the manipulation of these logical form representations.

Proponents of this hypothesis will say that, when a thought is tokened *only* in its logical form, it will be unconscious. It becomes conscious when, using a mechanism similar to the one Carruthers' has put forward (cf. 2.2.2), it is also tokened in natural language and processed through the mechanism of inner speech actions rehearsal—where a "full-blown phonological representation" (Botterill & Carruthers, 1999, p. 223) is also triggered. Modules, or System 1 processes, are posited by Botterill & Carruthers to be able to process these "natural language representations (of LF) as input" (Botterill & Carruthers, 1999, p. 224) making mentalese a superfluous theoretical construct. This is not without problems as we will see in 2.2.5

### 2.2.4 Language as transforming the mind

This last view is the most radical according to Botterill & Carruthers' (1999) discussion. According to Dennett (1991, 1994, 2009), language is a complete game changer. Before language evolved, brains—the hardware—were only running parallel “distributed connectionist processors” (Botterill & Carruthers, 1999, p. 216), and language acted as an operating system, a virtual machine, making it possible to run new kinds of software. This virtual machine, or Joycean machine as Dennett (1996) calls it, is entirely based on natural language. Using Dawkins’ (2006) meme idea, Dennett suggests that words, connections between them, ways of thinking about problems, and so on are memes running on this virtual machine. Language makes it possible, then, to “discover a new logical space, where you get the sorts of different behaviors, the sorts of new powers, the sorts of new problems” (Dennett, 1998, p. 130) that we associate with the kind of minds human beings have.

Words, understood as memes, are parasites that infect language-ready brains and modify them in a way that makes possible certain innovations we associate with human minds. Brains, equipped with language, went from devices that would make it possible to “[learn] new tricks” but only gave that creature a “short attention span” and no ability to plan for novel “long-term projects” (Dennett, 1991, p. 189) to devices making it possible to plan the construction of cathedrals (Dennett, 2009). I will say more on Dennett’s view and the digitization of information for minds, an idea also developed by de Sousa (2007, sec. 3.4), in 2.3.3.

I want to end this description of Dennett’s view with two notes—worries really, before criticizing a distinction proposed by Botterill & Carruthers (1999). First, while Botterill & Carruthers (1999) make clear that they endorse many claims made from the three other perspectives, their treatment of Dennett is short and they go straight to refuting it—this is why I am not going into more details here. It is not that these objections are not right or that they do not raise worries for

Dennett's framework[8]. However, the view is dismissed very quickly and, as I will develop in 2.3.3, there seems to be more to it. Botterill & Carruthers (1999) do not take it sufficiently seriously. The second worry is that they associate Dennett's view with Bickerton's (1995), citing Bickerton's claim that cognitive powers of prelinguistic hominids were very limited. Yes, language is the evolutionary building block that made it possible to think 'offline', but Bickerton's view is closer in spirit to the one developed in 2.2.1—at least in its most recent iterations. In his latest book, Bickerton makes clear that he does not endorse a Joycean view:

> It would be all too easy to summarize the take-home message of this book as "Advanced human cognition results from language." That is inaccurate even as shorthand. "Without language, advanced human cognition could not have existed." is better but still inadequate. It could be rephrased as "Human ancestors began to communicate with displaced reference, and that was what triggered the processes that eventually led to advanced cognition." But that, if more accurate, is unlikely to catch on as a slogan. (Bickerton, 2014, p. 263)

Therefore, in the present context, I will keep Bickerton's—whose views I will not be discussing further in the present context, but whose position seems to be closer to strong variants of the scaffolding view discussed above—and Dennett's views in different categories despite Botterill & Carruthers' grouping.

---

8. Although I *do think* that these objections are not right. The main objection they mention has to do with an underestimation of the mental powers of other species and nonlinguistic humans. Dennett could easily reply that the problem is rather that Botterill & Carruthers (1999) vastly underestimate what brains, these massive, parallel, and distributed connectionist networks can achieve. The other objection is related to concerns about the architecture of mind. They argue that Dennett's view is "inconsistent with the sort of central-process modularism defended [previously]" (Botterill & Carruthers, 1999, p. 217). This seems to be begging the question. Someone who would think Botterill & Carruthers are right would agree, but someone who thinks they are wrong about this kind of architecture or someone who thinks Dennett is right would not be convinced by this argument. I am not saying the Dennettian framework has no problems or limitations; I am saying that these two objections do not seem very threatening in the form they take in this passage.

### 2.2.5 Explicit and conscious thought

In 2.2.2 and in 2.2.3, I described two conceptions of language that Botterill & Carruthers (1999) distinguish. I want to suggest here that these two views are really a single one, viz. the conception of language as being the vehicle for explicit thought and the conception of language as being the vehicle for conscious thought are, in the end, describing two sides of a same coin. This is hardly surprising since, in the first place, Botterill & Carruthers (1999) understand these distinctions as being stronger or weaker forms of one another. I believe, however, that the distinction does not warrant listing them as different views. The main reason is that these two views do not differ sufficiently.

The main difference between the views discussed in 2.2.2 and in 2.2.3 has to do with the status of nonconscious thoughts and whether logical form (Chomsky, 1995) or mentalese (Fodor, 1975) is the most appropriate framework to make sense of the observed phenomena. This is certainly difficult to assess, especially since this is more or less given as a dilemma. The views, however, do not seem to be mutually exclusive.

Viger (2005) suggests, among other things, that sometimes natural languages can be more expressive than the language of thought and, through learning a given language, the language of thought can become richer. If this approach is right, then mentalese would gain in expressive power when we learn some natural language expressions—Viger is particularly interested in logical terms since these furnishes the combinatorial resources for the system. As Viger states it:

> By learning the logical/formal terms of a natural language we not only acquire the ability to entertain the logical concepts themselves, but all concepts that have logical structure, thereby greatly augmenting our minds. (Viger, 2005, p. 319)

This is a view that cuts across the positions described in 2.2.2 and in 2.2.3 as, in this case, the logical form changes unconscious processes but adding to the operations carried out in mentalese, operations that would be otherwise unavailable. This is not allowed in either framework. For the consciousness view, unconscious processes are not supposed to be modified by natural language and, for the explicit view, the existence of mentalese is rejected.

In the end, these three hypotheses might be wrong, or any one of these might be right. However, the existence of a position that cuts across the proposed distinction supports the idea that Botterill & Carruthers' (1999) distinction is not sufficient to capture the solution space. It seems more plausible that these various positions share a single solution space where researchers can adopt slightly different positions on how mentalese interacts with cognitive processes, beyond the possibilities entertained by Botterill & Carruthers (1999).

## 2.3 Remapping the landscape for the roles of language in cognition

The crucial benefit of Botterill & Carruther's (1999) framework is that it identifies some of the main positions in a vast, and sometimes messy, literature. Picking out these three[9] roles for language has paved the way for thinking more clearly about the identification of the role of language in cognition.

9. Given the argument in 2.2.5.

Building on this early work, Carruthers (2012) presents a new analysis of the literature where he argues more clearly for a point mentioned in his book with Botterill. He suggests that the views vary from one extreme to another where, on the one hand, language is seen as having no role in cognition and, on the other hand, it is seen as wholly determining cognition. The position according to which thought is completely independent from language might not be defended by anyone—if only because we interact with other people using language and this can influence one's thoughts. The other extreme, that thought is "conceptually dependent upon language" (Carruthers, 2012, p. 383) is not discussed here, or by Carruthers, mostly because—in both cases—we believe it to be discredited from the point of view of cognitive science (but see Andrews (2002) for a critique of Davidson from this point of view[10]). The other positions, including the ones discussed above, are thus different positions to be placed between these two extreme positions. Although I agree, as I discuss in 2.3.1, that the perspective offered by Whorf and the perspective offered by Vygotsky are different in terms of degrees, I believe that the two other positions are of a different kind. I will say more on this topic in 2.3.2 and in 2.3.3. My account will differ from Carruthers' (2012) in two ways: first, I keep Dennett's view as one of the interesting options while he dismisses it (cf. discussion in 2.2.4); second, I think the idea of an axis with these two extremes applies well to compare Whorf's and Vygotsky's views, but I do not think it works well to situate Carruthers' own view and Dennett's.

10. Andrews (2002) suggests that Davidson's view of the theory of mind is wrong because some people with autism have difficulties interpreting other people's minds despite using language perfectly well. Andrews (personal communication) mentioned that, presented with this phenomenon, Davidson told her that it was impossible that such people exist.

### 2.3.1 Language as a cognitive tool & contemporary Whorfian views

In 2.2.1, I discussed the diversity of positions under the "scaffolding" heading. I want to continue this discussion here, and insist on some of its most important features. The common feature of these theories is that they see language as a tool being used to change cognition in some ways. However, a lot here depends on how strong the claim is made to be. As we saw above, with Seuren's (2013) definition of the Whorfian thesis, the claim can be made very strong, such as the idea that language *determines* thoughts, or very weak, where the idea is rather that there is a slight, or a big, influence of language when it comes to thoughts. While the stronger version is rightly dismissed (McWhorter, 2014), the weak Whorfian thesis, together with Vygotsky's idea, allow the idea that language is a cognitive tool to be a powerful one, well supported in contemporary cognitive science research.

Colombetti (2009) offers a very interesting perspective on how language interacts with feelings. She develops a framework in which language can be used to give rise to new feelings, modify existing ones, or specify them (dividing a feeling in subclasses by using different words). She makes numerous parallels with the specification of senses, such as taste—she claims that oenologists can learn to make new distinctions and tell more subtle differences, and that they learn to do so through language-use. This fits nicely with the scaffolding view. Language is used, in various ways, with emotions that arise, e.g., by labeling these emotions or by introducing distinctions between different ways of feeling a given emotion, and can give a new flavor to the emotions in question, sometimes transforming the initial feeling by making it possible to focus on some aspects of the initial emotion.

The case reviewed above of the Pirahã people and the cognitive effect of their lack of specific numerals is also interesting here. In this case, it is not that they are unable to compare sizes of sets, but rather that they lack the tools we usually use in order to make these comparisons from memory. Recent evidence also has shown similar effects for remembering smells (Majid & Burenhult, 2014). In their experiment, Majid & Burenhult compared participants speaking English to the Jahai people. In color identification and naming, both groups performed similarly. However, in a different task, they have shown that the Jahai participants were very good at naming, identifying and remembering odors. English participants usually rely on loose comparisons: "It smells like x." Jahai participants, however, have a much more comprehensive lexicon for odors—they "have a lexicon of over a dozen verbs of olfaction that are used to describe a wide array of odors" (Majid & Burenhult, 2014, p. 267). Such results, I think, show that Colombetti's (2009) perspective is not only interesting to understand how language interacts with feelings, but also with other cognitive capacities, such as remembering quantities of objects and remembering and comparing odors.

In these cases, language becomes a useful tool for working memory. Frank and colleagues (2008), as I mentioned, see language as a device useful to compress information in a way that makes it easier to handle. This notion is close to Miller's (1956) notion of "chunking". Chunking is our ability to divide what we have to recall into groups, or chunks, in order to recall them in an easier way, and is a good example of how working memory can be expanded with proper use of linguistic abilities. It is easier, for example, to remember a 10-digits telephone numbers by chunking it into three groups than to remember the 10 numbers separately. Introducing labels and associating them with sensory experience makes it easier to recall the experience in question—making this a form of chunking by using linguistic labels to remember more effectively the richness of a past sensory experience. Giving precise names to odors—rather than relying on general similarities between them like English speakers according to Majid & Burenhult's (2014) data—makes it easier to remember them and compare them, helping the Jahai participants to do much better when doing these tasks from memory. Similar effects have been shown in music perception[11], where associating musical patterns to 'familiar folk-tunes' greatly helped novice participants to identify musical intervals (such as minor versus major thirds). The use of a label, it seems, makes it easier for participants to recall intervals and helps them use this information appropriately to compare a given interval with other stimuli (Smith, Kemler Nelson, Grohskopf, & Appleton, 1994).

Boroditsky & Prinz (2008) take stock of such phenomena and argue that perceptual information and linguistic information are combined to overcome limitations of either kind of process. What is interesting in their view—and this is a point supported by the other authors mentioned in this section—is the dynamics between the two streams of information and how they can combine, how they inform and enrich one another. Odors and emotions, for instance, are more complicated than what words make them, and the variations we can see between various languages support this idea. New words, then, could help extract—make explicit—new kinds of information from sensory input. As they write:

11. Thanks to John Paul Minda for this suggestion.

> Neither perceptual information alone, nor the sets of correspondences between elements in language alone, are likely to be able to amount to the sophistication, scale, and flexibility of the human conceptual system. [...] Combining information from these two input streams, as well as extracting the wealth of information that exists in the correspondences across input streams, can help overcome the shortcomings of relying on any single information stream and can reveal information not available in any one stream. (Boroditsky & Prinz, 2008, p. 112)

What these experiments do not show, however, is that cognition is altered to an extreme point where, e.g., someone speaking one language could not understand someone speaking another language. Specific languages seem to have limited effects, such as making it easier to think about an idea—creating a more direct link between the natural language and the mentalese token Fodor (1975) could say—but giving rise to entirely new ideas, otherwise inaccessible does not fall under this view of language as a cognitive tool. In extreme cases, language can modify the neural architecture of the brain (Donald, 2002), but mostly in helping create more and better connections between ideas from different domains.

An analogy offered by Bloom (1998) can be useful here. He gives it in order to help us distinguish the possibility that language is a very good and useful tool for some tasks, a tool that might even help perform tasks that are otherwise very hard, or impossible to do, from the possibility that "language explains people's ability to understand or generate this information in the first place" (Bloom, 1998, p. 215). He suggests that vision, as a capacity that makes it easy to access certain types of information, can make it very easy to access information that would be hard to access for someone who is blind (given that infrastructures are rarely designed with accessibility in mind). Many books (not available in braille or as audiobooks), maps, many aspects of the Internet, etc. are not accessible to them. Bloom insists that this does not mean they are less cognitively capable; a blind person only has more obstacles when it comes to access certain types of information in our culture that depends heavily on visual information. Language, understood as a cognitive tool, is just like vision in this analogy: it helps in accessing certain types of information more easily. Given this view, then, language can enhance some cognitive processes, but it is unlikely (although not impossible in more radical forms of Whorfianism) that it completely transforms how a process works.

The view of language as a cognitive tool, as a scaffold, makes two main claims: first, language gives better and easier access to some ideas, and it can help various cognitive processes perform their tasks better—working memory being an important example; second, language can guide the acquisition and development of ideas that would be otherwise very hard to access. The first claim is accepted by both those who support the weak and those who support the strong interpretations of Whorf's thesis. The second claim is only accepted by proponents of the strong version of the hypothesis. In both cases, however, this kind of view claims that language *enhances* various cognitive processes, and claims of a stronger effect of language on cognition from this point of view are usually discredited, viz. it is usually held that English speakers could in principle learn the Jahai vocabulary for odors and become, with time and training, as good as them at identifying odors from memory[12].

12. Surprisingly, perhaps, this remains to be verified empirically, as far as my knowledge of the literature goes.

### 2.3.2 Language and intermodular communication

As I explained in 2.2.2, language is what makes consciousness possible for Carruthers (2006, 2013); it emerges through multiple iterations of System 1 processes—it is the means through which these processes do most of their, mostly internal, communication. What changes in cognition according to this perspective is that language makes it possible to combine not only concepts, but all kinds of cognitive processes. Boeckx (2010) sees human cognition as being “full of various tools”, modules Carruthers would say, that are combined through language “into a flexible all-in-one tool that makes available a variety of solutions (tools) whose effects can be combined spontaneously” (Boeckx, 2010, p. 131). From this perspective, language enhances processes in various ways, but in a different way than what was presented in 2.3.1—in this case, mostly by giving these processes access to much more resources. This can be done in two ways from the perspective of the language of thought[13]: natural language can create shortcuts for mentalese but, in a non-Fodorian perspective, could also increase the expressive power of mentalese, making it possible to entertain new thoughts (cf. Viger, 2005). In this latter case, one could argue, we would go beyond simply an enhancement, something I will develop in the next section.

This kind of framework, as I mentioned above, is useful to think about how different levels of processes might interact. If Carruthers is right that System 1 processes are related to the processes of System 2 in the way he posits, language opens up new kinds of possibilities and makes it possible to adapt behavior in a precise manner to different contexts. Gomila (2012) suggests that this is exactly

13. This is not required for the view, but I believe it is compatible with it as I suggested in 2.2.5.

what dual-process theories bring to the table. The nice aspect of Gomila's proposal, however, is that he sees this in a dynamical framework rather than a simple causal one. It is not a simple procedure that happens at once. Language, throughout development, influences various aspects of cognition. Because of these "symbolic means", he argues, the basic functions of cognition "become transformed, both representationally and procedurally" (Gomila, 2012, p. 119) in order to augment the flexibility and complexity of thought.

It is essential to note, however, that this is not only a more radical version of the view that language is a cognitive tool, or a view that gives language a closer connection with thought. Language, in this view, is intertwined with thought. More than a tool, it becomes a very important part of cognition and its organization. According to this view, language is part of the architecture rather than merely part of its scaffold—to continue this apt analogy. This argument can be taken even one step further as I will discuss in the next section: language *can be* part of the building itself, not just the scaffold.

### 2.3.3 The rewiring thesis

Dennett does not deny that creatures without language have concepts, but in a very different way than language-users do. He writes that "[c]oncepts are things in our world because we have language" (Dennett, 1993, p. 546), and this is the case because of the way we possess concepts, viz. we can "*consider* [a] concept" (Dennett, 1993, p. 546), and think about what makes it what it is. This is relatively trivial, of course, but this is the tip of what Dennett considers to be the role of language in cognition. In 2.2.4, I developed the view, advocated by Dennett, that brains are massively distributed connectionist systems and they are transformed when language comes in since it takes over—it installs a virtual machine that becomes the brain's new operating system. This changes the brain and does so even if language is temporarily taken offline (Dennett, 1994), as it is suggested by Dennett's response to cases presented by Donald (1991, Chapter 3).

In order to keep this brief, I will not detail these examples, but Donald (1991) discusses a handful of cases of human beings not having language—cases of "brains without language". What is interesting about the cases he brings up is that, without linguistic abilities, at least not *online* linguistic abilities (e.g., cases of temporary aphasia), these human beings are still able to realize many of the things we think of as being tied to linguistic abilities:

> Episodic memory continues to function, skills are retained, general knowledge of the environment remains in effect, and the individual is able to cope with complex social situations. (Donald, 1991, p. 166)

Dennett (1994) disagrees on *why* these abilities exist or, in the cases of temporary aphasia, are preserved. He argues that this is because the brains of these people with temporary aphasia normally have language and this changed the way their brain is organized, even in the temporary absence of language. So, even during aphasic spells, these patients benefit from the transformative role of language. Even when inner speech is not available, says Dennett, language already shaped the mind in a way that allows these behaviors. We cannot say, however, that these brains are "without language" as Donald (1991) claims.

An example of the type of transformation language brings to cognition, and this transformation in Dennett's framework would be as pervasive as any other impact of language in cognition, is the ability to pick out particulars within a more general class. For de Sousa (2007, sec. 3.4), this modifies the way cognition operates. Other animal species, in a sense, react to generalities, in opposition to particularities which are—if this idea is right—only accessible to minds that have language. Nonlinguistic minds, from this perspective, have thoughts that are "*essentially general*" (de Sousa, 2007, p. 72), viz. thoughts are about one element of a general class of objects, and the reaction would have been the same with any other element of this class of object, and not one of its particular members.

This does not mean that nonlinguistic beings think about general objects (e.g., categories) *as being* general objects—abstraction also necessitates language, but rather that, within a given set, it is easy for linguistic creatures to react to one particular object within a larger set of similar objects, and take it *as being* this one particular object within this larger set of objects. This is the ability that makes it possible to attribute proper names to one object within a larger set. For de Sousa, a creature does not only need language, but it needs to "have a mastery of language sufficiently rich to distinguish a proper name from a common noun" (de Sousa, 2007, p. 75). In this way, de Sousa argues, language makes it possible to think about things that would not be accessible otherwise.

Dennett and de Sousa would most likely disagree on the consequences this has for the notion of intentionality. For de Sousa, full-fledged intentionality requires this ability, given by language, to identify particulars. In contrast, Dennett's view, the intentional stance, does not make space for such a distinction since "intentionality" only has a heuristic value for him. Nevertheless, this idea by de Sousa that language changes the way we think about objects suggests, like Dennett's view, that human beings have concepts in ways that are not available to nonlinguistic creatures. Whether or not this is really the case remains an empirical question, i.e., whether or not de Sousa is right to write that this is a feature unique to linguistic creatures, but there is a sense in which language gives access to thinking about individuals *as being these* individuals in a way that is not accessible without linguistic abilities.

Under this third view, then, language might sometimes serve as a scaffold—Dennett writes that words, without being understood, can become familiar and "it is these anchors of familiarity that could give a label an independent identity within the system" (Dennett, 1996, p. 150). However, language does more than this in that it gives access to new ways of representing information; it generally reorganizes the system. The scaffold is still useful to build the mind, but the very material the architecture of mind is constructed with, what *makes* the mind what it is, if the rewiring thesis is right, is language itself.

## 2.4 A hypothesis for the *multiple* roles of language in cognition

First, I must note that the distinctions made above are mostly to facilitate how we think about various roles language can have in cognition. They are meant as a proposal that should in no way be static. There might be more options than the three I have settled on here but I believe this to be a good start. Second, I wish to highlight my main worry with many of the approaches I mentioned: they attempt to explain all of cognition. The views I mentioned so far have in common this attempt to attribute this one role to cognition, and they sometimes do so by showing how their own framework, x, can explain the data explained by framework y but in a manner in which x offers a better 'theory of everything-cognitive' than y.

I think this approach is not the right attitude to take, and I defend here a pluralistic perspective for the roles of language in cognition. My qualm is that, by attempting to cover what other theories cover, views might do two things: first, they might weaken their own perspective to make the data well explained by the other theory fit in a more harmonious way (e.g., without having to claim exceptions) within their own; second, they might discredit some data or some of its interpretations on the basis that it does not fit well with this very good theory that happens to explain *a lot* of cognitive processes, but maybe not this one weird case. There is nothing wrong with these two strategies in and by themselves, but I believe they are counterproductive where the sole goal is to confirm a given theory about a hypothesized (sole) role of language in cognition.

Of course, having a ‘theory of everything-cognitive’ is desirable, but we currently do not have much evidence that makes it possible to anticipate what such a theory would look like. When it comes to the role of language, it seems dangerous to commit to the view that, in this theory, language will only have one assigned role. Other options have to be explored first. This is especially important since I think that we do not have good reasons to believe that language will have the same effect on all cognitive processes. I will attempt to show there already exist examples in the literature suggesting we should not assign a single role to language, and parsimony alone does not seem a sufficiently good reason to posit a single role for language in all of cognition. Moreover, I claim that the views I summarized above each present an aspect of the very complex interactions language—or to be more precise, parts of the faculty of language—will have with various cognitive processes or, to borrow the terminology from the architecture of mind literature, how the faculty of language might change different modules in different ways. This point about modularity is something we can take from the discussion of the global workspace perspective, even though this is not a consequence of it. Various perspectives on modular architectures are interesting, but one of Carruthers' (Carruthers, 2006, Chapter 1) points is that not all modules are the same, and not all modules are encapsulated in the same way. Given this is central to Carruthers' theory, it would be surprising to see him resist this point about the various degrees and various manners in which language could interact with these modular processes. The idea, then, would be that Dennett's rewiring thesis might be right, but it might be only so for specific processes—some modules are rewired because of their interactions with language—and other processes might be better explained using the first or the second approach identified in 2.3[14].

Given the three roles for language discussed above, i.e. as a tool, as a means of internal communication and as a transforming, rewiring agent, I believe we can identify categories of interactions between language and cognition[15] and that making such a division will help researchers think differently about assumptions they make. The idea is to offer categories of relationships between language and cognitive processes depending on if and how a cognitive process is modified by language. This classification should make easier the work on this topic. These categories are to be understood at a different level than the above descriptions. Here, I develop some ideas on how language interacts with cognition while, in 2.3,

---

14. This is not an idea rejected by Dennett. However, what he has written so far on this topic has been vague. It could be interpreted either as a commitment to the view that language changes all of cognition (some sweeping claims seem to suggest this), an idea that does not seem plausible, or as the idea that language changes different parts of the mind in different ways, the idea I am defending here.

15. These four categories are loosely inspired by Colombetti's (2009) distinction between "reporting and describing", "enhancing", "clarifying and constituting", and "making accessible". In the first case, language does not change the emotion; in the second and in the third cases, language changes the emotion in different ways; and in the fourth case, the emotion is only possible through language. The categories below follow this progression.

I was concerned about the roles that are assigned to language. This distinction is subtle but important; every category of interactions explored here cuts across the roles identified above. In other words, both the rewiring thesis and the idea that language is a cognitive tool will help us understand Category 2 and Category 3 as I will make clear shortly.

First, Category 1 is one I did not talk about much in these pages: the processes that do not interact with language. This category captures a lot of processes, but likely candidates might be modules understood in Fodor's (1983) sense.

Category 1

> These are the cognitive processes that would be present even if language were not present. These are, for example, the processes shared in an almost identical form with nonhuman animals, but there might also be some of these that are uniquely human. Some processes that appear very early in a child's development and do not change (or do not change much) afterwards could qualify to be in this category. They will usually appear before language and a lot of these capacities might be used by language once it is in place. Language plays no role, or only an indirect role, for processes in Category 1.

The second category is meant to illustrate the kind of processes that are only slightly modified by language. We have good indication (cf. the discussion of Colombetti's (2009) work in 2.3.1) that many of the senses are of this type. They are not modified in how they operate, but language can help with neighboring abilities that changes, e.g., how memory interacts with the process in question such as in the discussion of the work by Frank and colleagues (2008) and by Majid & Burenhult (2014), as well as the example from Allen (2014).

Category 2

> These cognitive processes are present in nonhuman animals but language enhances them or how they operate. This second category, then, regroups the cognitive processes that are potentially enhanced by their interactions with the language faculty. The use of language makes these processes perform better.

The third category includes processes that are more transformed by language. A good way of thinking about the third category is to take Dennett's view but apply it only to a single process. Language can change how certain parts of the mind work.

Category 3

> The third category includes the cognitive processes *modified* by language. These processes are present in a different form in a nonlinguistic creature, and language modifies, sometimes beyond recognition, how the process works. Retracing and reconstructing the origin of this process is a matter of comparative psychology. These processes might coexist alongside their nonlinguistic forms—implicit learning does not stop when explicit learning starts, but it might very well use the same sort of cognitive mechanisms[16]. Both can also interact with one another. It is not an enhancement in the sense that one of the processes does not just perform better or worse; its very way of functioning is changed, and it can be changed more or less radically.

16. It might turn out that "explicit learning" and "implicit learning" are processes much more complex than we initially thought—we might have to divide these into subprocesses as well. Thanks to Genoveva Martí for this point.

Learning is an example of this third category: explicit learning has some advantages, but also some disadvantages over implicit forms of learning. A study by Dreisbach, Goschke and Haider (2007) uses a task where rules are implicitly learned and applied with a great deal of success, but where explicitly learning the same rules impedes rather than improves performance. The kind of learning that is done explicitly rather than implicitly is also much faster, but potentially less durable as the literature on expertise seems to suggest (e.g., Dreyfus & Dreyfus, 1988, Chapter 1; Kahneman & Klein, 2009). System 2 processes, in Carruthers' framework, would be in Category 3. Empirically, this would mean that there might be different ways to investigate these different processes, as it might be the case with research on categorization (cf. Chapter 4; Piccinini, 2011; Poirier & Beaulac, 2011).

Lastly, some processes might not be well understood as being tied to cognitive processes found in other species. My hypothesis is that these would be relatively rare, but they remain a theoretical possibility.

### Category 4

The fourth category includes cognitive processes (most) human beings have that *require* language to exist. Any cognitive process that uses complex social institutions can qualify here, but there might also be other examples where language plays a crucial and necessary role that would render this process impossible to realize otherwise[17]. Some obvious examples are the sort of means of communication we use (it sounds trivial to say that communication requires

17. Maybe a finer distinction would be warranted, viz. one between how language itself modifies cognition and how social institutions supported by linguistic and symbolic abilities modify cognition.

language, but this is not necessarily so—it is necessary, however, for the *kind* of communication most members of the human species partake in), but also probably some sophisticated forms of predictions and anticipations we are able to make.

Yet, the importance of this fourth category is, I think, exaggerated. It is the one we usually think about when thinking about processes that would most likely belong to Category 2 or Category 3. For example, many theorists underestimate the cognitive capacities of nonhuman species in a way that makes them classify processes that are either enhanced or modified by language as being processes that do not rely on apparatus shared with nonhuman species. In the end, I see these categories as being a more general perspective on language and how it relates to cognition. The views analyzed 2.3 are rather theoretical perspectives on this question, but they each cut across these categories. For instance, the scaffolding view explains well various processes that would fall in Category 2 and in Category 3, and the rewiring thesis is mostly useful to understand processes that would be classified in Category 3 and in Category 4.

## 2.5 Conclusion

This analysis of Botterill & Carruthers' (1999) proposal, and my suggestions for amendments was meant as a first overview of a vast, but still sparse literature. Few philosophers have addressed directly and in detail the question of the role of language in cognition. The proposed division of the different positions, my analysis of how they relate to one another, and the suggestion that there might be more than one role at once for language in cognition, possibly depending on the process we are looking at are, as far as I know, new proposals that have the potential to enrich the debate and to create new discussions.

Looking at the role language has in a given architecture is only one other part of a proper analysis, but sadly one that has long been neglected. I believe that if we start looking into proposed frameworks—such as theories of the architecture of mind (modularity, dual-process, etc.), paying attention to assumptions in experimental protocols in many areas of cognitive science, and questioning the way we investigate language and minds from a philosophical perspective, we will discover new problems and might be able to propose new perspectives. This has the potential to offer a fresh look into a problem that has been an ancient worry of philosophers: what makes our species unique? Language might not be the answer, but studying it and its relationships with the rest of cognition and the way it is organized should bring the debate further along.

## References

Allen, C. (2014). Symbols as Scaffolding. In L. R. Caporael, J. R. Griesemer, & W. C. Wimsatt (Eds.), *Developing Scaffolds in Evolution, Culture, and Cognition* (pp. 231–247). Cambridge (MA): MIT Press.

Andrews, K. (2002). Interpreting autism: A critique of Davidson on thought and language. *Philosophical Psychology*, *15*(3), 317–332. doi:10.1080/09515080210000061111

Baars, B. J. (1997). *In the theater of consciousness : the workspace of the mind.* New York: Oxford University Press.

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. doi:10.1016/j.tics.2012.12.002

Bickerton, D. (1995). *Language and human behavior*. Seattle: University of Washington Press.

Bickerton, D. (2014). *More than nature needs: language, mind, and evolution.* Cambridge (MA): Harvard University Press.

Bloom, P. (1998). Some Issues in the Evolution of Language and Thought. In D. D. Cummins & C. Allen (Eds.), *The Evolution of Mind* (pp. 204–223). New York: Oxford University Press.

Boeckx, C. (2010). *Language in cognition: uncovering mental structures and the rules behind them*. Malden (MA): Wiley-Blackwell.

Botterill, G., & Carruthers, P. (1999). *The philosophy of psychology*. New York: Cambridge University Press.

Carruthers, P. (2006). *The architecture of the mind : massive modularity and the flexibility of thought*. New York: Oxford University Press.

Carruthers, P. (2009). An architecture for dual reasoning. In J. S. B. T. Evans & K. Frankish (Eds.), *In Two Minds. Dual Processes and Beyond* (pp. 109–127). New York: Oxford University Press.

Carruthers, P. (2012). Language in Cognition. In E. Margolis, R. Samuels, & S. P. Stich (Eds.), *The Oxford Handbook of Philosophy of Cognitive Science* (pp. 382–401). New York: Oxford University Press.

Carruthers, P. (2013a). Animal minds are real, (distinctively) human minds are not. *American Philosophical Quarterly, 50*, 233–247.

Carruthers, P. (2013b). The fragmentation of reasoning. In P. Quintanilla (Ed.), *La coevolución de mente y lenguaje: Ontogénesis y filogénesis*. Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú. Retrieved from http://faculty.philosophy.umd.edu/pcarruthers/The%20Fragmentation%20of%20Reasoning.pdf

Chomsky, N. (1995). *The minimalist program*. Cambridge (MA): MIT Press.

Chomsky, N. (2013). What Kind of Creatures Are We? Lecture I: What Is Language? *The Journal of Philosophy*, *110*(12), 645–662.

Clark, A. (2008). *Supersizing the mind : embodiment, action, and cognitive extension*. New York: Oxford University Press.

Colombetti, G. (2009). What Language Does to Feelings. *Journal of Consciousness Studies*, *16*(9), 4–26.

Dawkins, R. (2006). *The selfish gene*. New York: Oxford University Press.

de Sousa, R. (2007). *Why think? Evolution and the rational mind*. New York: Oxford University Press.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

Dennett, D. C. (1994). Language and intelligence. In J. Khalfa (Ed.), *What is Intelligence?* (pp. 161–178). Cambridge: Cambridge University Press. Retrieved from http://ase.tufts.edu/cogstud/papers/rolelang.htm

Dennett, D. C. (1996). *Kinds of minds : toward an understanding of consciousness*. New York, NY: Basic Books.

Dennett, D. C. (1998). *Brainchildren: essays on designing minds*. Cambridge, Mass: MIT Press.

Dennett, D. C. (2009). The Cultural Evolution of Words and Other Thinking Tools. *Cold Spring Harbor Symposia on Quantitative Biology*. doi:10.1101/sqb.2009.74.008

Donald, M. (1991). *Origins of the modern mind : three stages in the evolution of culture and cognition*. Cambridge (MA): Harvard University Press.

Donald, M. (2002). *A mind so rare : the evolution of human consciousness*. New York: Norton.

Dreisbach, G., Goschke, T., & Haider, H. (2007). The role of task rules and stimulus–response mappings in the task switching paradigm. *Psychological Research*, *71*(4), 383–392. doi:10.1007/s00426-005-0041-3

Dreyfus, H. L., & Dreyfus, S. E. (1988). *Mind over machine : the power of human intuition and expertise in the era of the computer*. New York: Free Press.

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255–278. doi:10.1146/annurev.psych.59.103006.093629

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, *8*(3), 223–241. doi:10.1177/1745691612460685

Everett, D. L. (2005). Cultural Constraints on Grammar and Cognition in Pirahã Another Look at the Design Features of Human Language. *Current Anthropology, 46*(4), 621–646. doi:10.1086/431525

Everett, D. L. (2012). *Language : The Cultural tool*. New York: Pantheon Books.

Fodor, J. A. (1975). *The language of thought*. Cambridge (MA): Harvard University Press.

Fodor, J. A. (1983). *The modularity of mind : an essay on faculty psychology*. Cambridge (MA): MIT Press.

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: evidence from Pirahã language and cognition. *Cognition*, *108*(3), 819–824. doi:10.1016/j.cognition.2008.04.007

Gawronski, B., & Creighton, L. A. (2013). Dual-Process Theories. In *The Oxford handbook of social cognition*. New York: Oxford University Press. Retrieved from http://publish.uwo.ca/~bgawrons/documents/GC_Oxford.pdf

Gomila, A. (2012). *Verbal minds: language and the architecture of cognition*. Boston: Elsevier.

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. doi:10.1126/science.298.5598.1569

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. doi:10.1037/a0016755

Keren, G., & Schul, Y. (2009). Two Is Not Always Better Than One: A Critical Evaluation of Two-System Theories. *Perspectives on Psychological Science*, *4*(6), 533–550. doi:10.1111/j.1745-6924.2009.01164.x

Majid, A., & Burenhult, N. (2014). Odors are expressible in language, as long as you speak the right language. *Cognition*, *130*(2), 266–270. doi:10.1016/j.cognition.2013.11.004

McWhorter, J. H. (2014). *The language hoax: why the world looks the same in any language*. New York: Oxford University Press.

Merleau-Ponty, M. (2001). *Psychologie et pédagogie de l'enfant : cours de Sorbonne, 1949-1952*. (J. Prunair, Ed.). Lagrasse: Verdier.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. doi:10.1037/h0043158

Piccinini, G. (2011). Two Kinds of Concept: Implicit and Explicit. *Dialogue*, *50*(1), 179–193. doi:10.1017/S0012217311000187

Poirier, P., & Beaulac, G. (2011). Le véritable retour des définitions. *Dialogue*, *50*(1), 153–164. doi:10.1017/S0012217311000163

Samuels, R. (2009). The magical number two, plus or minus: Dual-process theory as a theory of cognitive kinds. In J. S. B. T. Evans & K. Frankish (Eds.), *In Two Minds. Dual Processes and Beyond* (pp. 129–146). New York: Oxford University Press.

Seuren, P. A. M. (2013). *From Whorf to Montague: explorations in the theory of language*. New York: Oxford University Press.

Smith, J. D., Kemler Nelson, D. G., Grohskopf, L. A., & Appleton, T. (1994). What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition*, *52*(1), 23–54. doi:10.1016/0010-0277(94)90003-5

Tomasello, M. (2009). Universal grammar is dead. *Behavioral and Brain Sciences*, *32*(5), 470. doi:10.1017/S0140525X09990744

Tomasello, M. (2014). *A natural history of human thinking*. Cambridge (MA): Harvard University Press.

Viger, C. (2005). Learning to Think: A Response to the Language of Thought Argument for Innateness. *Mind and Language*, *20*(3), 313–325. doi:10.1111/j.0268-1064.2005.00287.x

Vygotsky, L. S. (2012). *Thought and language* (Rev. and expanded ed.). Cambridge (MA): MIT Press.

Whorf, B. L. (2012). *Language, thought, and reality: selected writings of Benjamin Lee Whorf* (2nd ed.). Cambridge (MA): MIT Press.

# Chapter 3 : Second Paper

*Penser, c'est aller d'erreur en erreur.*
*Rien n'est tout à fait vrai.*

Émile Chartier, dit Alain (2005)

# 3 Dual-Process Theories and Language

## 3.1 Introduction

Dual-process theories offer an ideal ground to illustrate the issues that come up when it comes to investigating the role language plays in cognition—in other words, the role of language in the architecture of mind and how it interacts with other cognitive processes[18]. My aim is not to defend the dual-process framework so much as to show that it offers a productive starting point for research. It allows for the identification of what I think are more general problems we find in this literature in cognitive science, especially when it comes to issues related to the architecture and the organization of cognition.

I focus here on how such problems can be framed within the approach. Other models of cognition[19] might end up playing an important role in drawing a more adequate picture of the architecture of mind. It can be very useful to look at different ways of modeling and at different modes of cognitive processing, and this plays a crucial role when thinking about the role language can have on cognition.

---

18. I will leave "language" undefined for the time being as it is not defined in most of the views I will be discussing. I will return to this concern in 3.4, where I will offer a characterization of what language is, and what kind of cognitive processes it refers to. Briefly, language is not *one* cognitive process, and distinguishing its many parts is an issue that should be of concern for dual-process theorists. In any case, we must keep in mind that giving an accurate, very precise, definition of language would need a paper of its own to fully explore as there is no consensus on the matter in linguistics or in philosophy.

19. These other models are, among others, one-process models but also tri- or quad-process models. There are also views of cognition that do not use divisions in systems or types of processes. Additionally, Gawronski & Creighton (2013) propose to distinguish between domain-specific dual-process models—such as a dual-process model for $x$, or a dual-process model within a given approach in psychology, formalized models (such as those found in modelization) and generalized accounts. I will only discuss this last category here.

However, before investigating how these other models of cognitive architecture differ from the dual-process brands—arguably the most influential for the time being, it can be interesting to see how dual-process theories themselves behave, if only in outlining their own defining characteristics. For many dual-process accounts, the contrast between linguistic and non-linguistic processes is interesting to look at—if only indirectly—in order to differentiate between types of processes. Also, many of these frameworks present this distinction as mapping onto other distinctions made in the dual-process literature; in this sense, conscious processes and linguistic processes would be the same in these so-called generalized accounts (cf. Gawronski & Creighton, 2013, pp. 295–301). In other words, in the generalized account, if a process is conscious it is somehow linked to language (and vice-versa).

I will first describe the generalized account of dual-process theories and criticize how the issue of language is handled therein. I will then look at variations on this initial account, mainly coming from recent contributions by Carruthers, and by Evans & Stanovich—the latter two because they are arguably the most influential in this literature, and the former because he proposes a unique account where language has a key role—and suggest that these frameworks do not do a much better job than competing theories at handling the role attributed to language. Third, I will propose that, in dual-process theories, for both Type 1 and Type 2 processes, researchers should be more careful when considering language as one of their defining features. If this is right, one other feature should be re-examined, and that is the "shared with animals" / "uniquely human" dimension.

## 3.2 The generalized accounts of dual-process theories

According to the generalized accounts of dual-process theories, there would be two systems, or two types of processes, at work in human cognition. According to the most common terminology, they should be labeled "System 1", or "Type 1 processes", and "System 2", or "Type 2 processes" (cf. Samuels (2009) for the difference between "Systems" and "Types" in the dual-process theory literature[20]). These two systems are dual in character, viz. they are described in opposition to one another. System 1 is automatic, fast, unconscious, its processes work in parallel (similar to Fodor's (1983) modules—Stanovich (2004) even explicitly refers to Fodor's view). It is also described as evolutionarily ancient, shared with animals and nonverbal. System 2 is then described as controlled, slow, conscious and serial. It is evolutionarily recent, unique to humans and its processes are linked to language. Importantly, these various characteristics are thought to co-vary: in many of these frameworks, System 2 processes will be conscious *and* linked to language, or even sometimes conscious *because* linked to language.

20. Briefly, cognitive processes are defined here, minimally, as the processes that manipulate, store and transform information. On the one hand, a *type* of process is a set of processes grouped together because this grouping is theoretically or explanatorily relevant—e.g., sharing a number of characteristics, maybe given shared causal mechanisms. In this latter case, the type of process would be a natural kind, but not all *types* of processes will be natural kinds.

Systems, on the other hand, rather refer to interconnected sets of components that give rise to a property or mechanism—the systemic property (Wimsatt, 1985). Systems will have many processes organized in a specific way and behave in certain ways. Committing to the existence of a system and its processes ("System 1 processes") is thus a stronger ontological commitment than a commitment to the existence of types of processes that may, or may not, be organized as a system and could form more than a single system ("Type 1 processes").

Still, we have to be careful when talking about such generalized accounts. Evans (2008) provides a general framework to think about dual-process accounts, but even in this review paper he identifies different threads, different ways to look at these theories. Notably, he distinguishes between approaches in social psychology and approaches used in the psychology of reasoning. Nonetheless, the family resemblance between these various theories is still striking. First, dual-process theorists tend to agree that minds are to be understood as being composed of two systems or two types of processes[21]. They also tend to agree that each of these types of processes will share a number of characteristics, and that these characteristics co-vary. Evans (2008) summarizes these characteristics using Table 1.

As we can see in this table, System 1 processes are understood to be modular, to function in parallel and to be evolutionarily ancient; we can see System 1 cognition as using mostly evolved heuristics embedded in modules that are adapted to execute some tasks in a given context. Sometimes, however, the response of System 1 processes is not adequate. When the heuristic is triggered outside its normal range or in a context that is, evolutionarily speaking, unfamiliar, it will likely misfire. This is because, according to the explanation offered in this literature, no optimization of a process is needed from an evolutionary standpoint: satisficing—performing in a good enough way in relevant contexts—will be enough.

21. The potential disagreements between a Systems or a Types view are not so much disagreements between subdisciplines (e.g., where the psychology or reasoning would argue for the Types view and social psychology would rather defend the Systems view) as they are related to how committed researchers are with regards to the duality of the Systems or Types and how strong their ontological engagement will be.

Table 1. Clusters of attributes associated with dual systems of thinking (adapted from Evans (2008, p. 257))

| System 1 | System 2 |
|---|---|
| Cluster 1 (Consciousness) | |
| Unconscious (preconscious) | Conscious |
| Implicit | Explicit |
| Automatic | Controlled |
| Low effort | High effort |
| Rapid | Slow |
| High capacity | Low capacity |
| Default process | Inhibitory |
| Holistic, perceptual | Analytic, reflective |
| Cluster 2 (Evolution) | |
| Evolutionarily old | Evolutionarily recent |
| Evolutionary rationality | Individual rationality |
| Shared with animals | Uniquely human |
| Nonverbal | Linked to language |
| Modular cognition | Fluid intelligence |
| Cluster 3 (Functional characteristics) | |
| Associative | Rule based |
| Domain specific | Domain general |
| Contextualized | Abstract |
| Pragmatic | Logical |
| Parallel | Sequential |
| Stereotypical | Egalitarian |
| Cluster 4 (Individual differences) | |
| Universal | Heritable |
| Independent of general intelligence | Linked to general intelligence |
| Independent of working memory | Limited by working memory |

This makes System 1 processes limited, and these limits are studied within the heuristics and biases research program. Kahneman (2011) presents this program in a way that makes clear that the goal was to find how these heuristics were sometimes contrary to what would be rational choices, viz. choices determined to be optimal by analytic System 2 processes. The important discovery of the heuristics and biases research program is that these cognitive biases—meant here only as a tendency to give a certain response in a given context—are not random. They follow patterns. This is because the modules use specific *rules of thumb* that will not vary according to different contextual cues (i.e., they are, among other things, encapsulated) in a manner that would be sensitive enough to avoid some biases (i.e., detrimental ones); only the slow, controlled processes can do so. For Kahneman & Frederick,

> [t]he persistence of such systematic errors in the intuitions of experts implied that their intuitive judgments may be governed by fundamentally different processes than the slower, more deliberate computations they had been trained to execute. (Kahneman & Frederick, 2005, p. 267)

Although many such theories exist in the literature, I am interested in those that attempt "to map various dual-process theories into a generic dual-system theory" (Evans, 2008, p. 256). Samuels (2009) explains that such accounts share two tenets: first, the distinction made in tables such as Table 1 will align. In this sense, as mentioned previously, "processes which exhibit one property from a column typically, though not invariably, possess the others" (Samuels, 2009, p. 131). Although the properties will align, *which* properties will be included in a given theory will vary as theories rarely commit to the whole list compiled by Evans (2008). We will see an extreme example of this with Evans & Stanovich (2013). The stress, then, is on which of these characteristics will be included since, according to the second tenet, the processes studied have to be part of either system or either type of process. There is no in-between[22]. A theory that would not commit to have a given characteristic be either in one or the other System or Type of process would exclude this characteristic from the inquiry. Stanovich's (2004) theory embodies these two tenets nicely. I will be examining his account in the

22. This partly explains why many variations on dual-process accounts attempt to introduce new types of processes—Evans and Stanovich, for example, both proposed different tri-process theories (Evans, 2009, pp. 46–50; Stanovich, 2009a).

next section, before criticizing, more generally, the dual-process accounts similar to his and then assessing critically the role attributed to language in such theories.

### 3.2.1 Stanovich's (2004) dual-process theory

System 1, in Stanovich's (2004) framework, is a *set* of systems—The Autonomous Set of Systems (TASS). He sees these systems as being numerous and having "their own triggering stimuli" without being "under the control of the analytic processing system [System 2]" (Stanovich, 2004, p. 37). He sees the System 1 processes as being modular in nature, but only retaining three key features from Fodor's (1983) account: System 1 processes are "fast, automatic and mandatory" (Stanovich, 2004, p. 40). Many System 1 processes also can operate in parallel. Although his view fits squarely within the dual-process approach, Stanovich thinks System 1 processes can handle higher level inputs and outputs (even though most of them do not; this is what Fodor dubbed the "shallowness" of inputs and outputs)—some of these higher level inputs might be linguistic, but their operation is still not associated with conscious experience. The *output* of System 1 processes can be conscious, viz. once the information is treated by the modular process, the result might be available to other processes including those of System 2.

System 2 has characteristics associated with the other side of Table 1: there is a single process that acts in a serial, controlled manner and its operation is general (i.e., it does not have a specific domain of operation). System 2 has to do with "central executive control, conscious awareness, capacity-demanding operations, and domain generality in the information recruited to aid computation" (Stanovich, 2004, pp. 44–45) and it "allows us to sustain the powerful context-free mechanisms of logical thought, inference, abstraction, planning, decision making, and cognitive control" (Stanovich, 2004, p. 47). How these System 2 processes are *realized* is what is of interest here.

Where does language come in? Stanovich does not include it in his main characterization of System 2, and it is not presented as being a central or a necessary feature, either of System 1 or of System 2 processes. But the kind of roles language plays in making it possible for System 2 to accomplish many of its defining features is interesting. First, "the analytic system is uniquely responsive to linguistic input" (Stanovich, 2004, p. 48), both internal (inner speech) and external. Inner speech acts as a mechanism for self-stimulation and makes it possible for some cognitive modules to have access to information they would not have access to otherwise. Language is also what introduces "more seriality into information-processing sequences in the brain" and is used to "forge new connections between isolated cognitive subsystems and memory locations" (Stanovich, 2004, pp. 48–49). By *more* seriality in this serial system, Stanovich might mean that language allows for better organization of actions and thought processes and that it helps with cognitive control—in this sense, language helps System 2, a serial system, to organize its sequences of actions and reprioritize goals. Language is also what makes it possible for an agent to build a narrative of his or her actions, such as in cases of post hoc rationalization. In addition to this, language makes it possible to learn new rules and apply them almost instantly—this is the trivial sense in which language facilitates learning in a social context—which is an important difference between System 1 and System 2. System 1 can adapt its rules but it takes a long time to do so (Kahneman & Klein, 2009), while System 2, upon hearing a new rule, can start using it right away.

In this description of the role of language, however, the water gets muddy because Stanovich mixes and matches elements from theories that, although not necessarily incompatible, claim very different functions for language in relation with cognition, especially the kind of processes he qualifies as System 2. In a sense, in Stanovich's model, language does everything and nothing—it is not stated as being *required* for many of the functions of System 2, but without language much of what System 2 is capable of doing is not possible. The worry is that although he appeals to theories such as Dennett's view of a virtual machine—where "analytic processing is carried out by a serial virtual machine that is simulated by the largely parallel brain hardware" (Stanovich, 2004, p. 48)—Stanovich does not seem to get on board fully with other implications of Dennett's view since he does not characterize language as being as central as it is the case in this "virtual machine" framework. For instance, on Dennett's view, language is essentially what makes the virtual machine possible in the first place: it *rewires* the brain. In other words, language changes the very way cognition functions—language has a shaping role that transforms the way the mind works (Dennett, 1994, 2009a).

A clear example of this central role for language in Stanovich's picture is the ability to think hypothetically, which requires the ability to form hypotheses about possible states of the world, and keep them in mind while decoupling the representations of what is real from these hypothesized states of the world. It is one of the main features of System 2, and largely what makes it possible for System 2 to correct mistakes of System 1 processes when they arise. On the one hand, Stanovich writes that decoupling "is *often* carried out by the serial, capacity-demanding analytic system" (Stanovich, 2004, p. 50, my emphasis). On the other hand, he later writes that "it *is* the analytic system that carries out the critical operation of decoupling representations from their anchoring in the world" (Stanovich, 2004, p. 99, my emphasis). Since he does not provide examples of decoupling carried out by System 1 processes—cases from animal cognition could be illuminating here, the latter affirmation seems to be the most plausible interpretation of how these operations are carried out. The role of language is, in this process, to

> [provide] the discrete representational medium that greatly enables hypotheticality to flourish as a culturally acquired mode of thought. For example, hypothetical thought involves representing assumptions, and linguistic forms such as conditionals provide a medium for such representations. (Stanovich, 2004, p. 50)

In theory, the possibility of carrying out such tasks without language is not left out, but—in this work at least—it is never defined or shown as plausible. This means that these analytic processes are not only facilitated by language but they seem to be conceptually tied to its presence, partly because they require domain general representations about supposed—but not real—states of world.

In other words, Stanovich says that System 2 processes are possible without language but the way these operations are conceived depends heavily on the presence and the use of language. This makes many System 2 processes not only linked to language—language seems to be *required* for many of these operations, or at least required for the operations as they are defined and presented—but it also makes these processes uniquely human. The problem with this, as we will see in the next section and in 3.4, is that it allows only for a very limited understanding of System 2 processes. These processes could instead be included in a wider, richer framework enabling a more precise characterization of many cognitive processes.

### 3.2.2 Criticisms of the generalized accounts

A description of cognition where System 1 or System 1-like processes alone explains all of animal cognitive capacities, and where System 2 is the only part of cognition that explains what is unique to human beings in the animal realm falls short of being satisfying when it comes to explaining how the mind works. The situation is likely much messier than what this type of model can allow for. Still, dual-process theories are an interesting step forward in research; the roles of both systems' processes are recognized and investigated in their own rights. They can be used as the basis of very useful explanations and descriptions of cognition. In this, I side with Kahneman (2011). Dual-process theories have a heuristic value although they might not be the most accurate way to characterize cognition. They can remain very useful for understanding some phenomena, but there are problems we cannot ignore as Keren & Schul (2009) point out.

One of the main problems identified by critics is that the attribution of characteristics to each "type" is constrained by a very rigid list—e.g., a process cannot be both automatic and conscious. In some accounts, the characteristics are construed as always co-varying[23]. As seen previously, one of these characteristics

---

23. It is essential to note here that this property of dual-process theories has been overemphasized. Stanovich, West & Toplak (2014) note that the purpose of tables such as Table 1

> was simply to bring together the many properties assigned to the two processes in the proliferation of dual-process theories of the 1990s. The list was not intended as a strict theoretical statement of necessary and defining features. (Stanovich et al., 2014, p. 80)

They add that the "main misuse of such tables is to treat them as strong statements about necessary co-occurring features—in short, to aid in the creation of a straw man" (Stanovich et al., 2014, p. 81). I will here avoid constructing such a straw man but will insist on implicit assumptions that are held within the work on dual-process accounts. The working assumption, as we saw with Samuels (2009), is that the characteristics co-vary, but the characteristics considered will be a subset of those in Table 1.

is that System 1 is nonverbal, and that System 2 is linked to language—we saw in details how this is presented in Stanovich's (2004) framework. In fact, many theorists go further and describe System 2 as being bound by language. The distinction is sometimes taken for granted. For instance, the descriptions are made in such a way that the processes that are nonverbal, evolutionarily ancient and shared with animals are those of System 1, and those that have anything to do with language are thought of as System 2. The claim rarely is so direct, but the description of what System 2 accomplishes focuses on the peculiarities it has in the human case without going into the details of what System 2 processes could do without language—this is the worry I detailed in 3.2.1. Regardless of claims to the contrary, or suggestion that this might not always be the case, System 2 is in effect characterized as being language-dependent.

Yet, there are good reasons to avoid presupposing that all of the abilities linked to language for human beings necessarily need language to exist. There are, I think, more refined and precise ways to assess and carve up the problem space of the role of language in cognition. There does not seem to be good reasons to take for granted that all of System 2 processes are uniquely human or that they are all necessarily linked to language (cf. Toates (2006)). This seems to be, however, the direction in which the uncritical acceptance of the role of language in cognition has led the literature. I see at least two reasons to resist such a view.

The first reason is evolutionary plausibility. Toates (2006) introduces a useful distinction between "uniquely human" and "uniquely developed in humans", that might have very important consequences for how we link dual-process theorizing with reflections on the role of language in cognition. In fact, it opens up a vista. A cognitive process or capacity, under such an understanding, would be uniquely human if we find no plausible precursor of the same kind in nonhuman species. Language is often given as an example since forms of animal communication are of a different *kind* than human linguistic communication (although this is not uncontroversial). "Uniquely developed in humans" rather means that we can find plausible precursors in nonhuman species; memory in human beings is different from the memory of bonobos or chimpanzees, but there are similarities suggesting continuity between the ways in which they operate. It might be because some of these views are not detailed enough and that the devil is in the detail. I will explore this possibility in 3.3 by focusing on two other frameworks in the dual-process theory literature.

The second reason to resist the view that System 2 is uniquely human and linked to language (and that System 1 is nonverbal) has to do with the explanation of certain simple cognitive phenomena, hard to explain under a standard view of dual-process theories. There are cases in the literature (e.g., Donald, 1991, Chapter 3; Lecours & Joanette, 1980) that show that very complex, reflective tasks can be done without active ("on-line") use of language. Moreover, if we take seriously Dennett's (1994) view that language has profound and lasting influence on (maybe only parts of) cognition, we can plausibly suppose that there is, at least, *something* verbal in System 1[24]. Karmiloff-Smith (1992) offers a great example of a framework in which we can conceptualize how some processes can *become* automatic. If we adopt Karmiloff-Smith's point of view, it is even possible that many capacities thought to be innate might be just like driving or speaking, viz. a skill that becomes

24. The literature on expertise suggests how some abilities initially under conscious control become automatic. Stanovich (2004) does not reject this possibility but does not detail it either; he acknowledges this is an option without offering much more to his readers.

automatic when rehearsed multiple times in the right circumstances. This idea of processes becoming automatic, or the very possibility of training intuitions, has also been explored by Kahneman & Klein (2009; see also Moors & De Houwer, 2006). Language understanding is also undeniably automatic and unreflective in most cases[25]; Fodor (1983) gives examples of understanding utterances of a known language as a clear example of a modular process. In 3.4, I will discuss how we might solve these kinds of worries.

## 3.3 Recent variations on the generalized accounts

The idea that language has a prime role in making possible System 2 operations is deeply entrenched in the dual-process theory literature. Evans explains that

> Language provides the means by which we can represent complex concepts, ideas, and suppositions in our minds, as well as communicate them to others. Indeed, it is *impossible to imagine* how the reflective mind [System 2] could operate without it. (Evans, 2010, p. 31, my emphasis)

In this sense, Stanovich (2004) certainly does not offer the most radical example of how language is intertwined with System 2 processes, but his characterization still makes it difficult to see how these processes would operate without language. Yet, the role of language seems unclear—it is incompletely explained how it can achieve such tasks. In two recent accounts, theorists have proposed ways of detailing the place language has in dual-process frameworks. I will, in turn, explore

---

25. Some lists of the characteristics attributed to System 1 and System 2 place language on the "System 1" side (and it is a module in Fodor's (1983) sense). Some processes are indeed associated with producing and understanding language, and they are automatic—grammar acquisition is also listed as a module by Stanovich (2004). The use of language is, in most cases, understood as being controlled however. This is why it usually lands on the "System 2" side. Whether language is on one side or the other of such lists, these characterizations leave aside key factors.

a framework Evans & Stanovich (2013) have put forward, joining forces, and Carruthers' (Carruthers, 2006, 2013a, 2013b) view since it details a very specific, and interesting, role for language in the architecture of mind, although not without issues, as I will make clear.

### 3.3.1 Evans & Stanovich's recent proposal

In a recent review paper, Evans & Stanovich (2013) address many concerns about dual-process theories and develop a bit further how they see the distinction between the two types of processes. Importantly, they recognize the misalignment of many attributes such as those listed in Table 1. In this new account, they consider only two dimensions to be of interest when distinguishing between Type 1 and Type 2 processes (cf. Table 2), but they also insist on the many differences between various accounts of dual-process theories. Even though they see other potential features as correlates, these two attributes are seen, in this new model, as being both necessary and defining.

Table 2. Defining features of Type 1 and Type 2 processes (adapted from Evans & Stanovich (2013, p. 225))

| Type 1 processes (intuitive) | Type 2 processes (reflective) |
|---|---|
| Defining features | |
| Does not require working memory | Requires working memory |
| Autonomous | Cognitive decoupling; mental simulation |

They still insist, however, that these are two *types* that we must distinguish and not, e.g., two extremes of a continuum. Type 2 processes, in this sense, are not a kind of variation of Type 1 processes; they have a distinctive purpose and act in a different way. Type 2 processes will be linked to "control states that regulate behavior at a high level of generality" (Evans & Stanovich, 2013, p. 230). They mention a lot of evidence for the relevance of such a distinction, the most convincing probably being that what is meant by Type 2 processes has high variability between individuals, and it is linked to general cognitive abilities. Type 1 processes are more or less the same across the board. In other words, the biases are the same for everyone, but some will be better at correcting biases. Stanovich has long argued that this kind of difference warrants a great part of the distinction between Type 1 and Type 2 processes.

This observation has led Stanovich and his colleagues to argue the following: the individual differences in using Type 2 processes to correct biases is correlated with measures of IQ and of fluid intelligence which, in turn, correlate with measures of working memory capacity. Type 1 processes are independent of such measures (IQ or working memory capacity), but an individual's performance in executing tasks that require some kind of decoupling or other form of correction of known cognitive biases is not. Type 2 processes are thus linked to working memory (Evans & Stanovich, 2013, pp. 235–236; Evans, 2008, pp. 270–271; Stanovich, 2009b, 2011). In contrast, the defining feature of Type 1 processes is their autonomy and the fact that their performance is not correlated with working memory capacity in an individual.

The importance of working memory here has to do, it seems, with one of the main features of Type 2 processes: their capacity for decoupling which requires a certain amount of cognitive resources to mentally simulate counterfactual scenarios. As we saw earlier, however, this decoupling ability is rarely discussed in a context where language does not come in, making decoupling conceptually linked to language-use or, more generally, to abilities made possible through language. In other recent contributions, Stanovich confirms the importance of language, which has only slightly changed since his 2004 account: "Language appears to be one mental tool that can aid this computationally expensive process." (Stanovich, 2011, p. 50) He then proceeds to restate some of the ideas mentioned in 3.2.1

He adds later on, while tracing connections between the literature on Theory of Mind (the ability to predict the behavior of others) and general accounts of Type 2 processes, that he follows "Mithen (1996) in thinking that language helped to break down an evolutionarily older modular mind by delivering nonsocial information to the social module" and Carruthers' view of language where it is "a mediating representational system that serves to integrate the modular mind" (Stanovich, 2011, p. 91). The vocabulary he uses to indicate possibilities or partial contributions (e.g., "can aid", "helped to") obscures a much more profound commitment to the kind of role language has in such architectures as I have suggested above. Language is not merely peripheral in Stanovich's characterization.

The issue here is not that language has an important role in itself—language is surely important in human cognition, but rather that this role for language is conflated with other kinds of processing. This, in turn, gives us a misleading account of the kind of processing we are interested in (mostly Type 2 processes in this case), but also an unclear account of just what language is supposed to be doing in this architecture. In 3.4, I will suggest ways to avoid encountering such problems. Before turning to this, since Stanovich mentions Carruthers' view, I will now detail how Carruthers offers an interesting variation on the dual-process theories literature where language has a very clear role. His view encounters a different kind of worry.

### 3.3.2 Carruthers' dual-process view

Carruthers (2006, 2012, sec. 6, 2013a, 2013b) puts forward his version of dual-process theory to contrast his view with some of those common in the growing literature on the topic. He identifies a number of problems within these influential accounts and offers what, I think, is a nice solution. It is not, as we will see, exempt of issues, but Carruthers' view is certainly a step forward.

He argues that the System 1 and System 2 distinction "should be abandoned" (Carruthers, 2013b, p. 1), although he believes there is a distinction between *intuitive* and *reflective* processes. He argues for this in a similar fashion to what we see in the Evans & Stanovich view detailed in the previous section, where there is a distinction between Type 1 and Type 2 processes. I will use here the Type 1 / Type 2 terminology used so far in this paper, rather than the one Carruthers uses, to facilitate comparison between his view and others[26].

More than rejecting the Systems view, Carruthers sees problems with the "default interventionism" view that dominates many dual-process accounts—including Evans & Stanovich's (2013). The default interventionism idea is that, by default, Type 1 processes will process information because they are usually much faster than Type 2 processes, in addition to not requiring much cognitive resources. To put it in a simplified way, following Kahneman (2011), brains are "lazy" and will take the shortest possible route to solve a problem—they will avoid any effort deemed unnecessary. According to the default interventionism perspective, Type 2

26. Carruthers (2013b) might resist this given some remarks in his paper, notably remarks about how features of each type of process are distributed. Given the discussion in 3.3.1, and Evans & Stanovich's (2013) proposal, I think he could be on board with this characterization.

processes need to override the activated Type 1 processes in order to replace the latter's faulty reasoning. Since Type 1 processes work well most of the time, they can be used by default, and Type 2 processes only intervene in problematic cases, given appropriate conditions (detecting the problem, having the right cognitive tools to correct it, having sufficient cognitive resources to succeed, etc.). This means that both kinds of processes can be activated at once, but not at the same time. Type 2 processes come into play only when they are needed, viz. when Type 1 processes make a mistake that is detected[27]. The problem with this, and this is where Carruthers' view shines, is that the two types of processes here seem to "exist *alongside* each other, competing for control of the person's behavior" (Carruthers, 2012, p. 395). From an evolutionary point of view, this seems a very costly solution; a whole new type of process, or in some dual-process accounts a whole new system, would have evolved to monitor responses of the processes already in place. Following Frankish (2004), Carruthers presents Type 2 processes as being "*realized* in those of System 1" (Carruthers, 2012, p. 395). This solution has the advantage of presenting the two types of processes as interrelated rather than independent from one another, and often competing for resources.

The properties of Type 2 processes, from this perspective, are the outcome of many cycles of Type 1 processes. The way this works is that we mentally rehearse actions, and these action rehearsals are broadcasted in a global workspace (following Baars, 1997) where they become available to other Type 1 processes. The input is

---

27. *How* it is detected is an interesting issue, but addressing it would bring us well beyond the scope of this paper. Stanovich (2009a, 2009b) has suggested a tri-process view of cognition to address this issue, but the jury is out to decide whether or not this is a satisfying solution. Evans (2009) suggested a competing view although he seems to have abandoned it since in favor of Stanovich's account (Evans & Stanovich, 2013).

processed by a given Type 1 process and its output is then ‘published’ on the blackboard, where it gets picked by another Type 1 process, and so on. I will detail shortly how this happens, but I first want to outline two advantages of his view.

The first advantage is that it gives us a good grasp on a possible continuity between the cognition of humans and other species, where the difference would be one of degree rather than a difference in kind (Carruthers, 2013a). While most Type 1 processes will be shared across the spectrum, the uniquely human features we observe would come from a difference in how many cycles the Type 1 processes can be handled. It also makes it possible that some animal species will have some degree of Type 2 processes—Carruthers thus addresses a concern he shares with Toates (2006). As Carruthers puts it: “other species of animal already possess the beginnings of System 2” (Carruthers, 2012, p. 395). If Type 1 processes are the only kind of cognitive process that are shared between humans and other species, some complex animal behavior becomes indeed very puzzling. The second advantage is that it explains clearly how Type 2 processes have developed in the way they have in the human species; their functioning relies on the evolution of systems for language production and comprehension[28].

The idea is powerful and very seductive, and the role for language is quite clear. Language is the interface that makes possible multiple iterations of Type 1

28. This is not the *only* difference between human and nonhuman animal cognition, but it seems to be the most important one. Among other differences, Carruthers lists: “an enhanced mindreading faculty, together with a drive to share mental states with other agents”, “a capacity for normative thinking and distinctively moral forms of motivation”, “greatly enhanced abilities for skill learning and fine-grained control of action” (Carruthers, 2013a, p. 243) and maybe more. It is interesting to note, however, that many of these differences do rely on the presence of language or, at least, powerful means of communication that would be akin to language.

processes in the global workspace, giving rise to Type 2 processes. Language, Carruthers argues, "led to a transformation in character of System 2" (Carruthers, 2012, p. 396). This transformation is that, with language production and comprehension, the array of action rehearsal available for global broadcast is larger: speech actions can now be rehearsed. In this picture, as Carruthers himself clearly states, "language plays an important constitutive role in distinctively human (System 2) thought processes" (Carruthers, 2012, p. 396). This idea is explained by de Sousa, who adopts a similar perspective on modularity and how language gives rise to new types of cognitive processes: "the information furnished by the diverse specialized modules are linked together by a universal, topic-neutral system of representation." (de Sousa, 2007, p. 81)

This elegant way of explaining Type 2 processes gives us a better grasp of some properties of these processes. The realization of Type 2 processes in cycles of Type 1 processes can explain why the former is slow in comparison and, because only one action can be mentally rehearsed at once, Carruthers has the resources to explain why Type 2 processes are serial. The actions selected can also be chosen, which gives us a way to grasp the controlled aspect of Type 2 processes.

This view is not without its issues however. First of all, it seems to take language to be *one* unified process, at the exception of the common distinction between language comprehension and language production, historically linked to the Wernicke and the Broca areas. *Many* Type 1 processes are surely needed to execute any speech action rehearsal. This does not have to be a problem for Carruthers but it needs to be detailed much more and, to do so, the language production and comprehension systems will surely have to be divided into more precise processes—a suggestion I will pursue in 3.4. The problem, however, might be more dire, and this is the second issue I want to raise: how can Type 1 processes, modules independent from one another, grasp the natural language content that is globally broadcasted in the workspace without having their own language comprehension apparatus? Making natural language the *lingua franca* of the mind is an interesting proposal but it seems to require a very heavy cognitive apparatus that, on its face, seems implausible[29]. Moreover, as I will develop in the next section, Carruthers' account of language and speech action rehearsal remain too nebulous to be operative in the way he suggests[30].

It is worth noting as well that, on this view, nonhuman minds are mostly composed of compartmentalized modules and that it is language that allows modules to exchange information. This claim has been challenged by Tomasello (2014, p. 130) on the basis of empirical inadequacy. He insists that so-called System 2 processes are much more common in greater apes than what Carruthers' account seems to allow for (Tomasello, 2014, Chapter 2). Another insight from the animal cognition literature comes from Camp, who has suggested—controversially—that baboons have cognition that is "plausibly like language" (Camp, 2009, p. 126) in the sense we are interested in, viz. how it structures certain kinds of modular outputs. Camp argues, for instance, that baboons can represent complex hierarchical dominance

---

29. A similar argument has been made by Davies (1998), pointing out that Carruthers relies on phonological form to be accurately represented in inner speech action rehearsal. Stainton (2006, pp. 177–190) raises more objections related to mine, but on his view on sub-sentential speech acts.

30. There are concerns as well with the proposed massively modular architecture he relies on (cf. Prinz, 2006).

structures[31] in ways that require the type of combination Carruthers says is permitted by language.

## 3.4 Type 2 processes and their links to language

One thing that should be of concern at this point is that the way the term "language" has been used so far is rather vague. In this section, I want to address this issue and discuss how this lack of precision might explain some of the issues I have been highlighting so far. I will suggest a way to understand "language" in these discussions of the architecture of cognition and, hopefully, pave the way for more research on this topic within this growing literature.

Language, it seems, is a very peripheral concern in many accounts of the dual-process theory within this literature. When it is not absent[32], its discussion lacks precision and the links to other issues are rarely explicit—the reconstructions of dual-process accounts I offered so far are, I think, representative of the state of the discussion. Language, however, should not be left out of the picture, and this is why I see Carruthers' (2006, 2012, sec. 6, 2013a, 2013b) view as a step forward. As I just mentioned, however, his own characterization of language also needs to be explicated.

Many recent proposals in this literature also remove "linked to language" (cf. Table 1) from the characterization of Type 2 processes. Although this might seem

31. Chapais (2010) has argued that Japanese macaques have similar cognitive abilities to represent matrilineal hierarchies.

32. To take one recent example, there is not much under the "language" entries in the index of Sherman, Gawronski, & Trope's (2014) anthology on dual-process theories in social psychology. The only serious discussion of the role of language has to do with concept grounding in one of its 38 chapters, covering most issues within the field.

to be on the right track, the way this is done is more likely a way to *avoid* the issue rather than to 'really' remove language from how Type 2 processes are understood and characterized. I think the discussion I offered in 3.2.1 illustrates this point well. Language should not be left out of the picture since we run the risk of characterizing some processes as having a feature *because* language is present while ignoring this very fact. Thus, when included in the discussion, the role of language must be made explicit in a way where its role, what it does in the architecture of cognition, is clear. Furthermore, if Type 2 processes are defined without having language as a core component, what Type 2 processes can do without language should also be made clear—e.g., what are their limits? Which Type 1 processes are involved in using language is another question of interest. How language interacts with both these types of processes is a third set of concerns that should, eventually, be addressed.

As the discussion of Stanovich's (2004; Evans & Stanovich, 2013) view made clear, I hope, language might be involved in a wide variety of cognitive tasks and intervene within many different processes. For some of these, language might be necessary—these processes would be impossible without it—but, for other tasks, language might merely enhance the abilities of some already existing processes. This is a possibility Stanovich implicitly entertains without committing to it. As we saw, the way he phrases the issue suggests that he shifts from one possibility—language is required for Type 2 processes—to another—language allows for better and more precise processing.

Making these concerns explicit will also help, I believe, in linking discussions of the architecture of the human mind to ongoing discussions in the animal cognition literature where many complex nonhuman species' abilities are investigated. How are they related to Type 2 processes? Using a more precise notion of what language is might be helpful to accomplish this—I will attempt a characterization of language before concluding this paper.

In this context, "language" refers to the *faculty* of language, defined as the set of cognitive tools that, together, make it possible to learn, to use (e.g., to produce sentences) and understand given languages—complex communication systems that have their own grammar (English, French, ASL, etc.). This account, of course, is far from perfect, but will suffice for now[33]. The faculty of language itself has been argued to include two subsets of cognitive apparatus: the Faculty of language—broad sense (FLB) and the Faculty of language—narrow sense (FLN). This suggestion by Hauser, Chomsky, & Fitch (2002), while controversial, can be useful in distinguishing the kind of roles language can have in cognition.

FLN is used to refer to a posited "abstract linguistic computational system" (Hauser et al., 2002, p. 1571). Recursion is the proposed candidate. FLN is the only part of the faculty of language unique to the human species. FLB refers to cognitive and sensory-motor characteristics that are required for language, many of which are shared with other species. Vocal cords and the lungs are examples of FLB, and so are many cognitive capacities without which language would not be possible—memory and various categorization processes being likely candidates. Fitch, Hauser, & Chomsky (2005) make clear that FLN might end up being empty—a proposal being that it might only be a unique arrangement of many FLB processes that give rise to the faculty of language. This does not mean that every part of

---

33. "Language" is rarely defined, even in discussions of its evolution. The idea of the "faculty of language" is what is of interest for our current purposes.

FLB will be found *as is* in other species, but the difference will be one of degree. For example, memory processes might rely on different cues or encode information in different manners in different species.

This framework can open many doors for discussions of the architecture of cognition and, in the current case, dual-process theories. The main take-home message should be that language is not monolithic and involves many processes, many of them being Type 1 processes. For instance, Chomsky's famous "language acquisition device" can be, under most characterizations, characterized as being a Type 1 process, since it operates automatically and unconsciously. But when it is said that language allows some processes to be "uniquely developed in human beings" (Evans & Stanovich, 2013, p. 236), it would be interesting to know which parts of the faculty of language are at play, which are required and how they modify a given process. Is a component unique in the human species necessary for a given Type 2 process? If yes, it means that a component that falls within FLN is being posited, and it might be crucial to understand language.

If this is an interesting direction for further exploration, Stanovich's (2004) view would gain a lot of traction if the role of language was made more precise. When he writes that "[l]anguage [...] introduced more seriality", that language is used "to forge new connections between isolated cognitive subsystems" (Stanovich, 2004, pp. 48–49), and so on, making clear which aspects of the faculty of language does what would make his characterization much more convincing and his explanation of these cognitive processes much more powerful. This kind of precision could also contribute to solving some of the issues Carruthers' view encounters. This would allow us to characterize how language performs the role it is attributed as well as making clear what aspects of which process is modified. Likewise, it would help to discuss issues related to aspects of human cognition that are unique to our species. Whether or not FLN is involved in a given process and whether or not a particular combination of FLB that seems uniquely human is necessary for a process to be carried out should be an essential part of this discussion.

Many Type 2 processes could be understood as strongly linked to language, while others might be less so, and still others not at all. Being able to distinguish each of these capacities might prove to be crucial. In such a framework, we can think of some Type 2 processes as nonverbal and of others as linked to language because this distinction is orthogonal to other similar distinctions (i.e., most of those in Table 1). Language will be involved in both Type 1 and Type 2 processes. Given that language has a wide variety of roles in dual-process accounts, this is not a discussion we should avoid.

Likewise, given that it is usually accepted within these frameworks that there are multiple Type 1 processes, corresponding more or less to a modularity of mind kind of view, each of these processes might be modified in different ways by language. There will not be *one* general role for language for Type 2 processes, but the same is true for Type 1 processes; each can be modified in different ways and each can, in the human case, have novel characteristics due to their interaction with one or more parts of FLB. None of the views we have reviewed here offer a compelling reason why this would not be the case. No evidence has been offered that language has a single role; there currently is no evidence that this role is the same throughout all cognition.

## 3.5 Concluding thoughts

In the end, I believe the dual-process framework is warranted, and some version of it should be defended. However, not all versions will fly. In this paper, I suggested that we should revise particular commitments made in how both types of processes are described, and sometimes defined. The accounts which end up posing more problems are those where descriptions of what Type 2 processes are doing are hard to distinguish from a general role for language processes—this has been the central issue I have been dealing with in this paper.

Of course, there are correlations between many features within a given type of process. When characterizing a type of process, however, as Evans & Stanovich (2013) make clear, not all of these features are to be understood as being on the same level. Some will be central, maybe even *defining*, and others will be peripheral or merely fortuitous correlates. These would be correlates that we do observe in most cases, but features that can be completely absent of a given Type 1 or a given Type 2 process—hence not features that we should focus our attention on.

In this paper, I defended the view that language is one of these features we should not focus our attention on. It is certainly a characteristic that seems central, even defining, to how a huge amount of processes we think of as being Type 2 processes work. However, responding to linguistic input is neither unique nor central to this type of process. Moreover, language is a complex phenomenon that involves many kinds of cognitive processes. I think that importing the FLN / FLB distinction from the literature on the evolution of the faculty of language would be most useful in better characterizing what is meant by language, which aspects we are talking about, and how they might be contributing to many features of human-specific (or developed uniquely in human) Type 2 processes. I think it will also be useful—but more empirical evidence is needed here—to better characterize nonhuman animals' cognitive processes that we might want to include in a broader account of what Type 2 processes are. Not only will this give us a better way to understand the architecture of cognition, but this will also provide us with more economical and precise explanations of how many cognitive processes work. This should give us better traction on this problem. In my view, this is a promising way to explain the seemingly huge cognitive gap between the cognition of human beings and that of other animal species. Although it might not settle the issue whether there is a difference of degree or a difference of kind between human and nonhuman cognition in the animal realm (Carruthers, 2013a; Penn, Holyoak, & Povinelli, 2008), assessing more carefully how language interacts within the architecture of mind would contribute greatly to this very discussion.

# References


Alain. (2005). *Propos sur l'éducation ; suivis de Pédagogie enfantine*. Paris: Presses universitaires de France.

Baars, B. J. (1997). *In the theater of consciousness : the workspace of the mind*. New York: Oxford University Press.

Camp, E. (2009). A language of baboon thought? In R. W. Lurz (Ed.), *The Philosophy of Animal Minds* (pp. 108–127). New York: Cambridge University Press.

Carruthers, P. (2006). *The architecture of the mind : massive modularity and the flexibility of thought*. New York: Oxford University Press.

Carruthers, P. (2012). Language in Cognition. In E. Margolis, R. Samuels, & S. P. Stich (Eds.), *The Oxford Handbook of Philosophy of Cognitive Science* (pp. 382–401). New York: Oxford University Press.

Carruthers, P. (2013a). Animal minds are real, (distinctively) human minds are not. *American Philosophical Quarterly, 50*, 233–247.

Carruthers, P. (2013b). The fragmentation of reasoning. In P. Quintanilla (Ed.), *La coevolución de mente y lenguaje: Ontogénesis y filogénesis*. Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú. Retrieved from http://faculty.philosophy.umd.edu/pcarruthers/The%20Fragmentation%20of%20Reasoning.pdf

Chapais, B. (2010). *Primeval kinship : how pair-bonding gave birth to human society*. Cambridge (MA): Harvard University Press.

Davies, M. (1998). Language, Thought, and the Language of Thought (Aunty's Own Argument Revisited). In P. Carruthers & J. Boucher (Eds.), *Language and thought. Interdisciplinary themes* (pp. 226–247). New York: Cambridge University Press.

de Sousa, R. (2007). *Why think? Evolution and the rational mind*. New York: Oxford University Press.

Dennett, D. C. (1994). Language and intelligence. In J. Khalfa (Ed.), *What is Intelligence?* Cambridge: Cambridge University Press. Retrieved from http://ase.tufts.edu/cogstud/papers/rolelang.htm

Dennett, D. C. (2009). The Cultural Evolution of Words and Other Thinking Tools. *Cold Spring Harbor Symposia on Quantitative Biology*. doi:10.1101/sqb.2009.74.008

Donald, M. (1991). *Origins of the modern mind : three stages in the evolution of culture and cognition*. Cambridge (MA): Harvard University Press.

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology, 59*(1), 255–278. doi:10.1146/annurev.psych.59.103006.093629

Evans, J. S. B. T. (2009). How many dual-process theories do we need? One, two, or many? In J. S. B. T. Evans & K. Frankish (Eds.), *In Two Minds. Dual Processes and Beyond* (pp. 33–54). New York: Oxford University Press.

Evans, J. S. B. T. (2010). *Thinking twice : Two minds in one brain*. New York: Oxford University Press.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science, 8*(3), 223–241. doi:10.1177/1745691612460685

Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition*, *97*(2), 179–210. doi:10.1016/j.cognition.2005.02.005

Fodor, J. A. (1983). *The modularity of mind : an essay on faculty psychology*. Cambridge (MA): MIT Press.

Frankish, K. (2004). *Mind and Supermind*. New York: Cambridge University Press.

Gawronski, B., & Creighton, L. A. (2013). Dual-Process Theories. In *The Oxford handbook of social cognition*. New York: Oxford University Press. Retrieved from http://publish.uwo.ca/~bgawrons/documents/GC_Oxford.pdf

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. doi:10.1126/science.298.5598.1569

Kahneman, D. (2011). *Thinking, fast and slow*. Toronto: Doubleday Canada.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 267–293). New York: Cambridge University Press.

Karmiloff-Smith, A. (1992). *Beyond modularity : a developmental perspective on cognitive science*. Cambridge (MA): MIT Press.

Keren, G., & Schul, Y. (2009). Two Is Not Always Better Than One: A Critical Evaluation of Two-System Theories. *Perspectives on Psychological Science, 4*(6), 533–550. doi:10.1111/j.1745-6924.2009.01164.x

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. doi:10.1037/a0016755

Lecours, A.-R., & Joanette, Y. (1980). Linguistic and other psychological aspects of paroxysmal aphasia. *Brain and Language, 10*(1), 1–23. doi:10.1016/0093-934X(80)90034-6

Mithen, S. J. (1996). *The singing Neanderthals : the origins of music, language, mind, and body*. Cambridge (MA): Harvard University Press.

Moors, A., & De Houwer, J. (2006). Automaticity: A Theoretical and Conceptual Analysis. *Psychological Bulletin*, *132*(2), 297–326. doi:10.1037/0033-2909.132.2.297

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31*(2), 109–130; discussion 130–178. doi:10.1017/S0140525X08003543

Prinz, J. J. (2006). Is the Mind Really Modular? In R. Stainton (Ed.), *Contemporary Debates in Cognitive Science* (pp. 37–56). New York: Blackwell.

Samuels, R. (2009). The magical number two, plus or minus: Dual-process theory as a theory of cognitive kinds. In J. S. B. T. Evans & K. Frankish (Eds.), *In Two Minds. Dual Processes and Beyond* (pp. 129–146). New York: Oxford University Press.

Sherman, J. W., Gawronski, B., & Trope, Y. (Eds.). (2014). *Dual-process theories of the social mind*. New York: Guilford Press.

Stainton, R. J. (2006). *Words and thoughts : subsentences, ellipsis, and the philosophy of language*. New York: Oxford University Press.

Stanovich, K. E. (2004). *The robot's rebellion : finding meaning in the age of Darwin*. Chicago: University of Chicago Press.

Stanovich, K. E. (2009a). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In J. S. B. T. Evans & K. Frankish (Eds.), *In Two Minds. Dual Processes and Beyond* (pp. 55–88). New York: Oxford University Press.

Stanovich, K. E. (2009b). *What intelligence tests miss : the psychology of rational thought*. New Haven: Yale University Press.

Stanovich, K. E. (2011). *Rationality and the reflective mind*. New York: Oxford University Press.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2014). Rationality, Intelligence, and the Defining Features of Type 1 and Type 2 Processing. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 80–91). New York: Guilford Press.

Toates, F. (2006). A model of the hierarchy of behaviour, cognition, and consciousness. *Consciousness and Cognition, 15*(1), 75–118. doi:10.1016/j.concog.2005.04.008

Tomasello, M. (2014). *A natural history of human thinking*. Cambridge (MA): Harvard University Press.

Wimsatt, W. C. (1985). Forms of Aggregativity. In A. Donagan, A. N. Perovich Jr., & M. V. Wedin (Eds.), *Human Nature and Natural Knowledge: Essays Presented to Marjorie Grene on the Occasion of Her Seventy-Fifth Birthday* (pp. 259–291). Dordrecht: Springer.

# Chapter 4 : Third paper

*Lento en mi sombra, la penumbra hueca*
*exploro con el báculo indeciso,*
*yo, que me figuraba el Paraíso*
*bajo la especie de una biblioteca.*

Jorge Luis Borges (1972)

# 4 "Concept" Heterogeneity and Definitions (Co-Authored with Pierre Poirier)[34]

## 4.1 Introduction

According to the Classical Theory of Concepts, concepts are definitions and definitions are necessary and sufficient conditions. The concept "bachelor" is encoded as the definition "unmarried man", viz. the concept of "bachelor" applies to someone if and only if he is male and unmarried. Psychological experiments[35] soon showed, however, that definitions were not the right kind of cognitive structure[36] to explain the performance of subjects and, in time, psychologists ended up rejecting any role for definitional structure in a theory of concepts. Participants, it seemed, use other cognitive processes and structures to categorize or distinguish objects, and psychologists spent the past 40 years designing clever ways to discover the nature of these processes and structures. In this paper, we argue that it was a mistake to completely reject definitions from concept science, a mistake that must be corrected if we are to properly understand this important part of human cognition.

34. The authors would like to thank Frédéric-I. Banville, Edouard Machery, Angela Mendelovici, John Paul Minda, Christopher D. Viger and Daniel A. Weiskopf for helpful comments and discussions on previous drafts of this paper. Thanks also to the audience for questions and comments at the 2012 meeting of the Society for Philosophy and Psychology in Boulder, Colorado—especially to Chad Gonnerman who offered a commentary on the paper. This paper is based on an argument initially published in Poirier & Beaulac (2011).

35. For instance by measuring the reaction times of subjects processing sentences whose concepts had different definitional complexity (for a standard review of the Classical Theory's problems, see Fodor, Garret, Walker & Parkes (1980); see also Pitt (1999) for a criticism of the Fodor et al. paper).

36. Following Machery (2009), we use the expression "body of information" to denote any cognitive structure that plays the role ascribed to concepts by cognitive scientists.

Our contention is not the back-to-the-future proposition that the Classical Theory was right after all and that concepts should henceforth be conceived as definitions. Our point is more subtle: the cognitive processes involved in categorization and discrimination tasks involve many types of cognitive structures, and changes in recent cognitive science allow us to find a restricted, but important, role in these processes for a type of structure closely akin to definitions as conceived by the Classical Theory. Because of the importance of this role, any complete explanation, mechanistic or not, of our ability to categorize and discriminate must include, we believe, such structures.

Two changes in cognitive science, one recent and still controversial, the other slowly building for the past two decades, constitute the background against which a new argument for definitions can now be offered: the rejection of the natural kind assumption for concepts (Machery, 2009) and the rise of dual-process theories of the mind (Evans & Frankish, 2009; Evans & Stanovich, 2013; Evans, 2008; Gawronski & Creighton, 2013; Kahneman, 2011; Stanovich, 2011). After a quick review of this new background for concept science, we sketch our argument for the (circumscribed) return of definitions, first by showing why, in light of this new background, definitions were rejected and second by explaining what role definitions play in human conceptual abilities. Our broader aim is to rehabilitate what we consider to be a fine notion, "definition", which has fallen in disgrace because of its close association with, in the minds of philosophers and cognitive scientists, a rejected theory. But the notion precedes whatever meaning it was given in the rejected theory, and cognitive scientists and philosophers are thus free to go back to other meanings of "definition" and redefine the notion anew in light of current knowledge in cognitive science. The situation, as we see it, is familiar in science: scientists reject one theory in favour of another; the description that was given to one of its central term is replaced by another, *but the natural kind referred to remains the same*. Just as physicists went from describing gold as "yellow malleable substance" to "the element with atomic number 79", cognitive scientists should now describe one of their natural kind, definitions, not as "necessary and sufficient conditions", but as we will propose below.

In the spirit of better understanding both philosophy and cognitive science, our hope is that this new framework will explain both why the common sense notion was recruited and given the meaning it was given by the Classical Theory of concepts and what it was that that classical theory got right about concepts (and, by extension, what it got wrong). Some challenges to this project will be raised in the last section of the paper.

## 4.2 A new landscape in concept science

Our argument for the circumscribed return of definitions in concept science rests on two conditions that specify the theoretical landscape in which such a return is possible: (1) the natural kind assumption about concepts is rejected and (2) it is agreed that some broadly dual-process view of the mind is accepted by all. This fact points to the way an opponent of the proposed return of definitions might argue against us. It also explains why the previous theoretical landscape in concept science (roughly the past forty years) was inimical to definitions, a point to which we shall return in this paper. For now, we simply present these two conditions.

### 4.2.1 Rejecting the natural kind assumption

Two main goals of science are prediction (of future events) and (adequate) representation (of the structure of the world). On a widespread view of science today (e.g., Giere, 2006; Ladyman & Ross, 2007), scientists can only achieve these goals once they have constructed scientific vocabularies (or typologies) that track *natural kinds*, viz. when the terms they use refer to categories whose members (objects, substances, events, etc.) possess stable sets of projectable properties. When this is achieved, scientists can reliably extrapolate to the whole category discoveries made about any one of its members. Natural kinds thus underwrite a disciplined practice of induction and prediction using law-like statements. A member of a scientific typology that does track a natural kind is called a "natural kind term" (Bird & Tobin, 2012; Schwartz, 1979)[37]. In this context, a "natural kind hypothesis" (regarding a term $T$ in a scientific typology) is the belief that T is a natural kind term. Accordingly, an important methodological task of any scientific discipline is checking the value of its natural kind hypotheses, thereby making sure its typology is made up of natural kind terms. Geologists, as is well-known, improved the quality of their predictions and theoretical representation of the world when they stopped using "jade" as natural kind term (i.e., when they rejected the natural kind hypothesis regarding the term "jade") to assume instead that the pair of terms "jadeite" and "nephrite" are natural kinds terms (Kim, 1992).

37. There are major debates regarding the existence, nature and function of natural kinds, and about the semantics of natural kind terms. We will not go into these debates here. We only require that there are natural kinds, that natural kind terms can refer to them and that science furthers its goals when its vocabularies are made up of natural kind terms.

When evaluating natural kind hypotheses in science, three things must be borne in mind. First, a given natural kind hypothesis may be well founded (there are many distinct compelling pieces of evidence supporting it), or it may simply be a working hypothesis, assumed correct until proven wrong. In the first case, the hypothesis is part and parcel of the discipline's knowledge. In the latter, its function is more heuristic, and we will follow Machery (2005, 2009) in flagging the status of such working hypotheses by naming them "natural kind *assumptions*". Of course, it may be the case, perhaps quite often in some disciplines, that the hypothesis lays somewhere between these poles: a working hypothesis backed by some *suggestive* evidence, but no more. Moreover, when a given term is only assumed to be a natural kind term, there is the question of whether there are alternative natural kind assumptions or hypotheses available. Second, as Griffiths (2004) remarks, some terms may denote "minimally" natural kinds, viz. categories where the generalization of discoveries made about some of its members is only better than chance. In such cases, the question facing those who evaluate natural kind hypotheses is not whether a given term refers to a natural kind or not, but whether a scientific discipline would be better served by a different natural kind term. Finally, as also observed by Griffiths (2004), kinds are natural (i.e., allow projection) only in the context of a given set of properties. The properties of pets are not projectable to other properties of interest to physiology or zoology, but they might be projectable to some set of properties relevant to social science or zootherapy. Accordingly, "pet" is not a natural kind term of anatomy or physiology, but it might be one of social science (e.g., theories about the demographics of a certain economic activity—the purchase and care of animal companions or decorative animals) or zootherapy (e.g., theories about the use of domesticated animals for therapy with seniors or sick children).

This last consideration means that categories are not natural kinds absolutely but only relative to some scientific theories. All these factors may be simultaneously active to complicate the evaluation of a given natural kind hypothesis: within a given scientific domain, a natural kind hypothesis may be well supported but only refers to a minimal natural kind whereas another may be supported only by suggestive evidence but for a much stronger natural kind.

One scientific domain where these complicating factors converge is psychology. As van Gelder (1995) argues, the Cartesian view of mind has beset modern psychology and its descendants (especially cognitive science and neuroscience) with a general *ontological homogeneity assumption*, according to which "one basic ontological story works for all mental entities" (van Gelder, 1995, p. 59). This general assumption makes natural kind assumptions at all levels of the conceptual hierarchy, from the very general likes of "mind" and "cognition" to the more specific likes of "remember" and "infer", the default assumption against which scientists must argue. It used to be assumed, for instance, that "memory" was a natural kind term, referring to one neurological system, but is now widely believed that there are multiple memory systems, that is, that the natural kinds terms to be considered in memory research are (at least) the more specific "working memory", "short-term memory", "long term-memory", and so on, all implemented at different levels of neurological organisation (from molecular to system-wide neurological). Similarly, it used to be assumed that the term "reasoning" refers to one type of mental activity, but many psychologists now argue that it sometimes refers to the activity of an evolutionarily ancient automatic system, and sometimes to the activity of a more recent and, as we will see, probably linguistically-based (or invaded) system.

One current proposal for such a restriction is particularly germane to our proposed rehabilitation of definitions in concept research. In his *Doing without Concepts*, Machery (2009) argues that many (if not most) psychologists of concepts have assumed that the central term of their discipline, the term “concept”, is a natural kind term. That is, psychologists took for granted that only one type of body of information is at work in concept experiments, and that the goal of their experiments is to inquire into the nature and function of this one single type of entity: what *it* is made of, what information *it* carries, what role *it* plays in cognition (recognition, reasoning, action, etc.), how *it* is acquired, how *it* reacts when faced with atypical instances of a category, etc. They have assumed that concepts “share scientifically relevant properties” (p. 54) and that, accordingly, the “class of concepts will yield numerous inductive generalizations (p. 54). In short, they have adopted the natural kind hypothesis regarding the defining term of their discipline.

Machery (2005, 2009) argues, however, that concept experiments put *many* types of bodies of information into play (prototypes, exemplars, and theories chief among them) and thus that the term “concept” in psychology refers not to one but to many natural kinds. Each kind may be solicited depending on the task at hand, but they may also be called on all at once by some cognitive processes: thinking about dogs may bring to mind prototypical information about dogs, one or a few exemplars of dogs and causal information about dogs (e.g., dogs bark at strangers). Machery thus rejects the idea that “concepts constitute a homogenous class about which specific, scientifically relevant generalizations can be formulated” (Machery, 2005, p. 449). According to Machery, prototypes, exemplars and theories are different kinds of bodies of information[38] functioning according to different principles[39], a view he calls “the heterogeneity hypothesis”. It is no surprise, thus, that the various accounts of concepts generated controversy: the three “competing paradigms” study different kinds of bodies of information. However, it must be added that the three share important features, and this makes concept science appear more unified than we believe it actually is: the activation of prototypes, exemplars and theories is typically fast, automatic and unconscious. But this unity, we will argue below, is an artifact of experimental procedure.

The natural kind assumption in concept science had two effects on the field. The first, as Machery notes, was to pit the various psychological-neurological accounts of the nature of concepts against one another. Proponents of the prototype view of concept had to defend their theory against that of proponents of the exemplar view and the theory view (and similarly for each of them). But this rivalry only makes sense if “concept” is a natural kind term. Rejecting the natural kind assumption means that each account focuses on a distinct natural kind, carved out from the original but now rejected “concept” category. The second effect, as we argue below, was to squeeze “definitions” out of concept research in psychology.

Machery’s suggestion has the major advantage of ending decades of counterproductive debates in philosophy and psychology. Questions such as “What is the correct account of "concept"?” now become hopeless. Rejection of the natural

---

38. In the case of prototypes, the information encoded is statistical knowledge about the typical properties of a class; in the case of exemplars, it is the properties of one or a few individuals of that class; and in the case of theories, it is knowledge about the relations of members of the class with the environment.

39. E.g., prototypes and exemplars are similarity-based (they are distinct because they compute similarity in distinct ways, but see Virtel and Piccinini (2010)), but not theories.

kind assumption modifies the landscape of concept research and opens the door for new approaches, new theories, and new accounts of the underlying cognitive architecture. One door that Machery's account opens is that the new kinds of concept science can be coreferential: there can be a prototype, an exemplar and a theory for a given category (or, indeed, many of each), and these may be instantiated and involved in parallel processes. One may now also think of competing and collaborating prototypes, exemplars and theories (explaining, e.g., slower or faster response time). And, importantly for us, Machery's account opens theoretical space for definitions in concept science; as we see it, his account shows that the rejection of definitions might have also been another fruitless consequence of the natural kind assumption. But to show this, we first need to introduce the second condition that sets up the new landscape for concept science, the adoption of a dual-process view of the mind.

Before we turn to this, however, there is one other point we must address here. Machery (2005, 2009) believes that rejection of the natural kind assumption motivates a form of "scientific eliminativism"; researchers interested in cognitive processing should eschew the concept of "concept" altogether. If Machery's rejection of the natural kind assumption has, we believe, put an end to fruitless debates in concept science, his scientific eliminativism, may have done the reverse; in the past few years, various accounts of concepts were proposed to square the continued use of "concept" in science with the rejection of the natural kind hypothesis. Our view follows from Machery's rejection of the natural kind assumption but remains neutral with respect to his scientific eliminativism. It is thus compatible with propositions that the concept of "concept" should be kept as a theoretical entity in the psychology of concepts. For example, Weiskopf's (2009) pluralist theory views concepts as superordinate structures constituted of distinct kinds of psychological structures. He believes that a category is represented by one (higher-level) concept that has many parts, including prototypes, exemplars, essential properties, words, etc. In this sense, he explicitly rejects Machery's scientific eliminativism (see Machery (2009, pp. 243–245) for a reply). These parts are stored in what he calls a long-term memory store (similar to Prinz's (2004) notion of a long-term memory network). When thinking about a given object, not all parts of the long-term memory store have to be activated in order to constitute a concept: a concept is constituted when some parts are retrieved from the store in working memory. Even though our account might challenge parts of what Weiskopf proposes (definitions might have, in such a picture, a different role than, e.g., prototypes and exemplars), we think that it is quite clear that it could be compatible with our account; so would Machery's proposal. Here, short of additional empirical evidence, we think we can remain neutral on the topic of this disagreement.

### 4.2.2 Adopting a dual-process view of the mind

Dual-process theories are now becoming widely used in cognitive science (cf. Evans & Frankish, 2009; Evans & Stanovich, 2013; Evans, 2008; Gawronski & Creighton, 2013; Gawronski, Sherman, & Trope, 2014). These theories posit that there are two types of cognitive processes, often labelled "Type 1 processes" and "Type 2 processes" (Evans, 2008; Samuels, 2009), whose characteristic properties are opposed (duals). "Type 1 processes" refers to the many parallel processes that are automatic, unconscious, fast, mandatory and, thus, used by default (in some architectures they are akin to Fodor's (1983) modules). "Type 2 processes", for its part, refers to cognitive processes that possess the converse characteristics, viz. serial, controlled / intentional, conscious and slow. Type 2 processes are usually associated with higher cognition (including thought, language use, etc.), but they are not always described as necessarily requiring language to work (Toates, 2006). This also does not mean that Type 2 processes use a representational format that would hinder any interaction between them and Type 1 processes. Following current literature, we will take it, for present purposes, that the Type 2 processes that concern the present argument are linked to language or some other form of explicit representation, but this is by no means necessary for our view[40]. Various dual-process theories differentiate themselves on a number of levels, including, but not limited to, the extent of the set of opposing properties posited essential to account for cognition, and the importance they assign to given properties in the set.

Although becoming ubiquitous in cognitive science, dual-process theories have also attracted much criticism (e.g., Gigerenzer & Regier, 1996; Keren & Schul, 2009; Machery, 2011) where they are (rightly) said to be oversimplifications (see also 3.2.2). For example, the strong oppositions posited between the processes taken to be duals simply do not hold in many cases (e.g., sequential processing may emerge from massively parallel computations) and many of the properties posited do not necessarily cluster in neat sets (e.g., some automatic processes are evolutionary ancient but not all are). Any retreat from neat clusters of opposing traits weakens the posited duality between "systems" or "types of processes" that is at the heart of such theories. Notwithstanding these very real problems with dual-process theories, we will adopt a *dual-process view* of the mind, that is, the view that the mind is

40. As we will explain below, as we understand them, Type 1 processes can also be linked to language and their outputs can be made explicit, which distinguishes our position from, among others, Piccinini's (Piccinini & Scott, 2006; Piccinini, 2011) view about how to distinguish between two kinds of concepts, implicit and explicit.

made up of (generally) two broad sets of (mostly) opposed processes, a view to be fleshed-out in a (yet to be devised) proper dual-process theory of cognition. We are not committed to any particular dual-process theory held in this literature; for our needs, it will suffice that a theory of the kind developed in the dual-process literature, or any successor theory that posits multiple processes (forms of tri- or quad- models would qualify), adequately characterizes the general architecture of cognition. We simply claim that some processes correspond to the Type 1 description, and that some correspond to the Type 2 description. This will help in distinguishing between the various processes at work in the concepts literature. Moreover, it will help to better understand, and to put into context what is meant when it said the some concepts are "used by default" (Machery, 2011).

We understand, like Kahneman (2011) does, these two so-called systems as "characters in a story" about how the mind works but, as with many stances we can use in science, the Type 1 / Type 2 distinction can help science move forward by making clear methodological issues in experiments (see below our criticisms of current methodologies in concept science). It is not our goal, however, to defend here this type of position. The reason we are interested in dual-process views of the mind is that we believe that categorization processes can be divided in two broad types, having dual properties, and that the two types of processes have an important role to play in humans' categorization processes. We also think that, because of the way concept science is generally approached, that one of these types of processes is poorly understood.

In the context of concept science, we believe that such "System 2" categorization mechanisms will have the following properties: they are used in a controlled manner (i.e., requiring attention), and their content will be transparent to the participant while being used (explicit). This might have the consequence that the ability to use these processes successfully will be correlated with working memory capacity, as predicted by recent accounts of dual-process theories (Evans & Stanovich, 2013), but we see this as more of a consequence of how these processes operate than as a cause.

From the get-go we need to qualify what we mean here: first, while these features, controllability and explicitness, might also hold true for other mechanisms studied by cognitive science, we do not believe that—and we are not committed to—the idea that these features of categorization processes can generalize to all of the other cognitive processes that could be classified under the "Type 2" label. Different processes within the "Type 2" category might have different characteristics. Second, we are not saying that every controlled or explicit process that occurs during a categorization task is of Type 2. When asked about their categorization procedures, participants are likely to make explicit aspects of Type 1 concepts that they use, since they are the most easily accessible to mind (they are *used by default* following Machery's definition of concepts in psychology). For example, although people are not aware of the exact processes at work when they compare an examplar to objects they categorize, the features they name when prompted to do so might be features of this exemplar as it is more readily accessible. We will say more on this topic in the next section.

## 4.3 Type 2 concepts

A man's children have been found dead and he admits stabbing them repeatedly, but the trial finds that he has not murdered them: the community is shocked. A logic student struggles to understand disjunctions as true when both disjuncts are.

A scientist must be especially careful not to overextend the consequences of her findings. One is surprised to learn that a young-looking woman, who is one of the best runners at her running club, is a grandmother. These are all situations, we believe, in which Type 1 and Type 2 concepts conflict. At the end of this section, we will come back to these cases and explain why we believe this. But, first, we need to say what we take Type 2 concepts to be.

We agree with Machery that prototypes, exemplars and theories are bodies of information that cognitive processes use by default[41] (Machery, 2009). Their activation is typical of Type 1 processes: they are fast, automatic and used by default, etc. Because of their central role in Type 1 processes, we shall call such bodies of information "Type 1 concepts". Our contention is threefold. First, in this section, we claim that some cognitive processes also use bodies of information that are typically *not* fast, *not* automatic and *not* used by default. Because they are almost exclusively used in Type 2 processes, we shall call such bodies of information "Type 2 concepts". For reasons we explain below (section 4.5), we believe that such bodies of information should be viewed as "definitions", despite the controversial history of that term in concept science.

Moreover, we claim (section 4.4) that the Type 2 processing of definitions influences the acquisition and processing of Type 1 concepts in a way that makes the study of definitions in concept science essential. If some cognitive processes use Type 2 concepts and if, moreover, Type 2 concepts influence Type 1 concepts, then restricting concept science to the study of Type 1 concepts is a problem: such an

41. Cf. 4.6.2 for a discussion of why we think that the processes posited in the theory theory approach are Type 1 processes.

endeavor can only provide a limited view of the processes it studies. Theories of Type 1 concepts do not focus either theoretically or experimentally on concepts we learn and apply explicitly and thus have little to say about their involvement in cognitive processes or on how they may contribute to the construction of Type 1 concepts. We will cover this topic in section 4.6.We take up each of the points that characterize our positive view in turn.

### 4.3.1 What are Type 2 concepts?

What happens when someone is asked how she knows something is a bird (Brooks & Hannah, 2006)? Typically, she will answer by giving a list of features that she takes to be diagnostic of birds, say, sings, has feathers and can fly. Similarly, to come back to a previous example, we could ask a person how she knows an action done by someone is a murder and, presumably, she would give us lists of other actions (e.g., holding the knife) or mental items (e.g., he was violent). These lists of diagnostic features are bodies of information; they are groups of meaningful items (e.g., "singing", "having feathers") that are relatively stable across time in the same individual (answers to the same question at t1 and t2 will be highly correlated) and across individuals (answers to the same question by two individuals will also be correlated). As philosophers and cognitive scientists have made clear, such lists of features cannot be used to pick out the extension of the English word "bird" (but more on this later). To explain what we take Type 2 concepts to be, we will in the following sections (1) give some examples of such lists, (2) characterize the type of situation in which these bodies of information may be processed explicitly and (3) explain what consequences this processing step has for such bodies of information. Our reasoning here is strongly influenced by the work of psychologist Lee R. Brooks (Brooks & Hannah, 2006; Brooks, LeBlanc, &

Norman, 2000; Brooks, Squire-Graydon, & Wood, 2007). Before going into these details, we want to raise some of the types of questions that are yet to be resolved by focusing on understanding such a body of information.

"Murder", everyone can agree, is *not* "causing to die" (Katz & Fodor, 1963). Does this mean that these bodies of information serve no cognitive function? That is, to stress the point once more[42], does the fact that such lists of items are *not* the linguistic meaning of the words or expressions mean that these lists serve no cognitive function whatsoever? If so, one might wish to ask why are such bodies of information brought to mind when asked whether one knows that something belongs to a certain category (e.g., bird, murder)? We take up the question of the cognitive function of these bodies of information in the next section. It is more important, however, to begin by characterizing them—as a first sketch we propose whether or not they are linguistic, structured and task-specific. Unfortunately, because definitions (Type 2 concepts) are all but excluded from the domain of concept science, there are few empirical studies we can draw on to characterize them. But many questions concerning Type 2 concepts are worthy of systematic empirical investigation.

The first is whether the items in the list must be *linguistic*. When the body of information is brought to working memory in order to answer a question asked by someone (How do you know something is a bird?), then of course the items brought to working memory will be linked to language. The point of the task is to produce a linguistic item (an utterance) in response to a linguistic item, so it is only natural

42. Presentations of this paper at various venues shows that we cannot stress enough the fact that our point is not to bring back the old story about the conceptual analysis of "murder" and such.

that language will be involved in the process. But is this necessary? To answer this, we need to get an idea of the various types of situations in which Type 2 concepts may be brought to working memory. If day-to-day question-asking provides one type of case, concept experiments provide another. In typical concept experiments, subjects may be asked to categorize a number of novel objects, greebles (Gauthier & Tarr, 1997) for instance. Some of the features that distinguish greebles have a ready-made lexical item to name them (e.g., "round") but some do not. When they wish to draw attention of subjects to such features experimenters may say: "Check for the little squiggle up there, just below the curvy line". To describe the feature, experimenters use language to describe their perceptual experience, which corresponds presumably to that of subjects.

The second cluster of questions is linked to the *structure* of these lists, viz. whether they are, or not, structureless. A list of items would possess structure if some items in the list were in some ways more important than others. Say we reproduce the situation described by Brooks & Hannah (2006) under various conditions: we ask the person to give us the single (or two, three, etc.) most important items for how someone knows that something is a bird. Is the first (second, third, etc.) item in all such lists correlated? Does the fact that she answers, say, "flies" in a one item list predict the first item in a two (three, etc.) item list? To our knowledge, such simple experiments, designed to show the structure (or not) within lists of diagnostic features individuals bring to mind when asked how they know a thing (event, etc.) belongs to a category (e.g., how she knows something is a bird), are yet to be done. Identifying variations where the tasks are done under time-pressure or with counter-examples at hand (e.g., for birds, showing pictures of penguins might have an effect of identifying "flight" as an important characteristic) or not could also be revealing.

A third set of questions would be whether such lists are *intelligent* or *task-specific*. Yarbus (1967) has shown that saccades are "intelligent" in the sense that saccades act as a function of the cognitive task one is asked to perform (e.g., recognizing, scanning, reading). A similar question could be asked of the lists of features brought to mind in order to answer questions: do they vary as a function of cognitive task or context? Will the same list (in the same order) be brought to mind when someone is asked how she knows something is a bird rather than when she's asked what feature is important when drawing a bird? Or when one has to set-up a classification of birds?

To fully understand the nature of Type 2 concepts, other questions could also be asked: do they have an ontogeny (are the characteristics brought to mind by infants the same as those brought to mind by adults and, if they differ, how are they different?)? Do these lists vary as a function of an individual's relevant specific knowledge? Will the three lists brought to mind by Joanna the Plumber be the same as the one Martha the Ornithologist or Gisèle the Avid Birdwatcher bring to mind? Answers to these questions (and many others) would give us a better understanding of Type 2 concepts, which would put us in a better situation to assess their role in human behaviour as well as understand their relation with the Type 1 concepts that psychologists have been studying for almost half a century now. Nevertheless, we believe we know enough about Type 2 concepts to take a first stab at answering these questions, which we will do in section 4.4 below.

No matter what characteristics Type 2 concepts turn out to have, the simple fact that lists of diagnostic features are brought to mind in some contexts is sufficient to explain the conflicts with which we opened this section: they are all situations where our Type 1 concepts (prototypes, exemplars or theories) tell us something falls into a category whereas the list of diagnostic features endorsed by a certain community tells us otherwise. In the first case, a legal definition of "murder" conflicts with our prototype of murder. In the second, the accepted definition of disjunction (i.e., inclusive disjunction) conflicts with prototypical use of "or" (i.e., exclusive). In the third, the scientist's operational definition of, e.g., "intelligence", may conflict with the essentialist theory-theoretical concept of intelligence possessed by most members of the community. Finally, in the last case, the formal definition of grandmother (mother of a mother) conflicts with our prototype of a grandmother (e.g., Betty White). Then again, we now need to make a convincing case that definitions are concepts in their own right and not, e.g., merely background knowledge.

## 4.4 The cognitive role of Type 2 concepts

Using an example from Brooks & Hannah (2006), we saw that a type of body of information is produced when subjects explicitly generate information (lists of features) regarding category membership when asked questions such as "How do you know something is a bird?" and we speculated about the properties these bodies of information might possess. Whatever properties such lists turn out to have, however, two things are clear at the outset regarding their status in concept science. First, they are not sufficient to pick out linguistic categories. The extension of the linguistic item "bird" cannot be given by looking for things that fly, sing and have feathers[43]. If, as was customary a generation or two ago, you define concepts as that which determines extension, then you will object to calling these bodies of information "concepts" (Type 1 or Type 2). It should be clear by our calling them "Type 2 concepts" that we do not take concepts to be in the extension-determining business. Second, it is also clear that if one believes that the term "concept" refers to one (and only one) natural kind whose nature it is the task of concept scientists to find out, that is, if one accepts the natural kind assumption for concepts, explicitly generated lists of features will not be a prime candidate for the referent of "concept". But assuming a framework in concept science where concepts are not in the extension-determining business and where "concept" is not assumed to be a natural kind term, the relevant question to ascertain the status of these lists of features in concept science is what function they serve in cognitive processes. This is the question we now turn to.

Our central claim to that effect is that Type 2 concepts play a role in the acquisition of Type 1 concepts (prototypes, exemplars and theories). Our claim, we should point out, is not that the cognitive processes that give rise to Type 1 concepts *need* Type 2 concepts. Evolutionary and ontogenic considerations suggest that Type 1 concepts can be formed in the absence of Type 2 concepts, and that they probably are in non-human animals and in very young human infants. Our point is that in an environment where lists of diagnostic features are exchanged between individuals (be it an elementary school, a science lab or a court of law), Type 2 concepts shape the content of Type 1 concepts. Here again, we will follow evidence from empirical work done by Brooks and colleagues (Brooks & Hannah, 2006;

---

43. Note however that you have to be pretty imaginative to imagine something that jointly possesses the three features but is *not* a bird (or a singer in a bird costume on a transatlantic flight).

Brooks et al., 2000) and the simple model of prototype formation offered by neural networks (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Maddox, 2005).

The first question to ask is why and in which circumstances lists of diagnostic features are produced. Brooks, LeBlanc & Norman (2000) show that, when medical students and experts are given specific diagnostic information about a patient's problem, the probability that they notice the relevant features in images increases. The relevant features should be obvious to participants since they are taken from medical textbooks, but the probability that they will in fact notice them nevertheless increases by 20% when the diagnostic information is given to them verbally. Moreover, the authors point out that subjects report seeing the diagnostic features more when the correct diagnostic is provided to them. When they see a bird, people see any number of its various features. When they are told its diagnostic features (e.g., flies, sings and has feathers), i.e., why it is a bird and not a mammal, the probability that they will notice these features when looking at birds increases. In other words, being told that something is a bird in relation with its relevant diagnostic features makes said diagnostic features more salient. This means that being given a Type 2 concept thus orients one's perceptual processes towards its diagnostic features. The same type of example becomes even more interesting in cases such as those discussed by Machery & Seppälä (2009): being told that "a tomato is a fruit" (a definition) increases the probability that people who know what a fruit is (i.e., those who know why something is a fruit and not a vegetable) will notice features that are specific of how tomatoes grow (which are relevant to their being a fruit) rather than of how they are used in the kitchen (which are not). *Mutatis mutandis* for whales: when someone is told that whales are mammals, and not fish, and given proper knowledge about what mammals are, people will notice more whales' mammalian properties (e.g., the absence of gills) than how similar they might be to fish (e.g., they live in the water).

In short, we saw that when producing classification rules, individuals produce Type 2 concepts (lists of features) and that Type 2 concepts make some features more perceptually salient to the individual than others. To see how these facts may be relevant to concept science generally, let's assume a simple feedforward, backpropagation trained neural network model of categorization. It has been shown repeatedly (see, e.g., McClelland & Rogers, 2004) that, when given vectors of input features corresponding to various exemplars of a category, such networks will generate prototypes that reflect the statistics of the given input features. Given a set of animal features and (as target) the animal they are diagnostic of, the network (properly structured and trained) will build a prototype for each animal that reflects the statistical distribution of diagnostic features among the animals. Now, it is often implicitly assumed that the input vectors to such neural networks correspond to direct input from the organism's environment, viz. that the contents of input vectors reflect the activity of some sensor. But the fact that the content of the input vectors are fully individuated features that can be labelled with words suggests that it is better to think of the input vectors as inputs to a late cognitive process, one that occurs after the first stages of perceptual processes have done their work. Thus, one could model Brooks and colleagues' (2000) results as showing that Type 2 concepts, by making diagnostic features more salient, affect the probability that such a feature will make it into the input vectors of prototype formation mechanisms, viz. Type 2 concepts affect the distribution statistics of inputs to prototype formation mechanisms, which will in return affect the affect the content of the system's prototypes.

If we distinguish two types of learning situations, the infant's and the novice's, the foregoing suggests the following picture of concept learning. Take the infant first. Their prototype formation mechanisms first construct prototypes that reflect the statistical distribution of features in the organism's environment. If most instances of *cathood* with which an infant interacts is Bubbles, the family fat brown cat, then the infant's cat prototype will weigh the family cat's brown color, bigger size and perhaps, behaviorally, its laziness and frequent eating more than other cat features. But at some point in his cognitive development, the infant will come into contact with other cats and will be given a Type 2 concept of cats, for instance that "cats are predators", making Bubbles' predator features more salient, thereby increasing the probability that such features will make it into the input vector of her prototype formation mechanisms (e.g., Bubbles' fangs and claws). If just thinking of Bubbles as a cat increases the probability that the features made salient by its "cats are predators" Type 2 Concept, then the child's cat prototype will be skewed towards the predator features of cats instead of Bubbles' plump appearance.

A similar process holds for atypical members of a category. Naïve observers will categorize dolphins as fish because exemplars of dolphins look like fish (e.g., tuna) and prototypical dolphins share salient properties with prototypical fish (swimming as mode of locomotion, water as where they live, and so on). Indeed, children will, at first, assume that dolphins act like most species of fish (Gelman & Markman, 1986). However, with more information, and often by being told by others that "dolphins are an aquatic mammal" (a definition), attention will focus on other dolphin features and behaviour; instead of attending to the mode of locomotion and the medium in which dolphin's live, features such as their need to breathe at the surface of the water and to breast-feed their young will be brought forth. In other words, within the theory theory framework, forcing naïve observers to focus on different and perhaps less prima facie salient properties, the conditional probabilities between events may change (more on this in section 4.6.2). The same goes, *mutadis mutandis*, for tomatoes (Type 1: vegetable-like; definition: fruit), penguins (Type 1: not bird-like; definition: bird flying underwater), whales (Type 1: fish-like; definition: mammal), and Hollywood-zombies (Type 1: alive looking; definition: not alive; cf. Machery & Seppälä (2009)).

On this view, the process of concept acquisition through Type-2 concepts is a "cognitive game changer". Once acquired, Type 2 concepts guide the acquisition of other, more automatic bodies of information (prototypes, exemplars, and theories). Cognitive processes involving prototypes, exemplars and theories are more adequate to account for participants' responses in concept experiments, and thus to provide an account of on-line individual cognition. But the responses subject give in such experiments may already reflect the shaping work Type 2 concepts had in their acquisition. Type 1 concept formation is a life-long, on-going process. As their Type 1 concepts are forming, individuals may be verbally corrected, or they may encounter situations where they must explicitly monitor and modify their responses (cf. Carey, 2009). These cognitive processes (understanding verbal corrections, self-monitoring of usage, and modification of inadequate response) illustrate the role we see for Type 2 concepts—sometimes even just having the label can have an impact as Carey's (2009) discussion of placeholders suggest.

Definitions, on this view, can have both a causal role (guiding learners to attend to some instances or some features of instances of a category at the expense of others) and a normative role (making sure that correct notions are acquired—this is especially true of course for institutionally controlled concepts as in science or the law). Rehabilitating definitions allows us to highlight the important relation between Type 1 and Type 2 concepts: both often cooperate by having reciprocal influences on each other. Type 1 concepts are the basis on which we acquire definitions—but sometimes this acquisition is triggered or modified by Type 2 concepts, via teaching (including reading or following a guide), showing, or experimenting under guidance. The way we intuitively parse and categorize a scene has a strong influence on how we linguistically represent and think about the world. Yet, Type 1 concepts sometimes have to be inhibited, or "taken offline", as when a jury must suppress automatically activating prototypes of murder to apply a legal definition. With overlearning, definitions might even sometimes come to be used by default, associated with the very meaning of a word, as Machery illustrates with "uncle" (brother of a parent): "*some* default bodies of knowledge might be definitions" (Machery, 2010, p. 433). We are skeptical that this can happen. We think it would be more plausible that using definitions can become almost automatic after multiple uses, but that definitions will mostly help other processes, such as those forming prototypes and exemplars, to identify the right kind of objects over time. In order to address this point, however, we have to say a bit more about methodology in 'mainstream' concept science.

## 4.5 Two projects for definitions

The classical theory of concepts thrusts definitions as its central theoretical construct and understands definitions as necessary and sufficient conditions. An important reason that explains this way of understanding concepts has to do, we believe, with the usual conflation between individual cognition and the scientific enterprise (i.e., psychology as philosophy of science writ small; see, e.g., Fodor (2000, p. 52)). Biased by this conflation, psychologists ask definitions to do two things at once: (1) account for individual understanding and use of categories—a project that falls within the domain of the psychology of concepts, but also (2) account for how scientific communities develop ways of carving nature that serve their goals of understanding and prediction—a project that falls (more) within the domains of the sociology and the philosophy of science.

In individual cognition, definitions will serve to explicitly correct prototypes, reject some exemplars in favor of others, or change the probabilities underlying a Bayesian network, as we outlined in the previous section. Moreover, as we mentioned previously, definitions can also serve to introduce a new category, opening a space (Carey (2009) calls this a "placeholder") that prototypes, exemplars and theories can then fill. Being told about the difference between two species of birds, an amateur bird watcher will explicitly bring to mind the distinguishing features of each species while bird watching. The same goes for elms and beeches. Exemplars of each species will thus be distinctly remembered and, with time, prototypes and even theories of the two species will develop.

In the scientific enterprise, definitions play a role in the social identification, use, and often challenge of those explicit characterizations of categories that are most useful to further science's goals of representation (adequate carving) and prediction. In astronomy, a planet is, since August 24th, 2006:

> A celestial body that: (a) is in orbit around the Sun, (b) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, and (c) has cleared the neighbourhood around its orbit. (IAU, resolution B5, online: <http://www.iau.org/static/resolutions/Resolution_GA26-5-6.pdf>)

Any object that does not satisfy all of these criteria is not a planet, and any that does is[44]. Such definitions are closer to necessary and sufficient because such conditions are in general well-suited for these purposes. The same goes for most scientific, legal and mathematical concepts, but also for many common categories that have explicit definitions (e.g., "your brother's stuff" is "everything on your brother's side of the room, except the computer").

Scientific definitions are different from those illustrated by the dolphin example above because the conventions established by scientists rest on more than simple individual understanding. Science is a social project that aims at producing knowledge with the highest form of justification and, to do so, controls its definitions publicly. Although scientific definition will, of course, influence how we individually categorize and make distinctions, both are nevertheless distinct; we would not expect a parent to use the scientific definition of planet above in everyday life, e.g., when correcting her child. The goal of an account of scientific definitions is not the explanation of individual cognition; concept experiments will not reveal anything relevant to what a scientific definition is or how it is built and regulated; philosophy of science might.

Indeed, we already said that definitions need not be necessary and sufficient conditions; that the latter are only one form of Type 2 concept. Conflating the psychological and scientific projects regarding concepts—as mentioned above when we discussed Fodor's idea that psychology is philosophy of science writ small—is most probably in large part responsible for their identification. By conceiving definitions as Type 2 processes, we link them to explicit, voluntary processing, and

44. Rumors are that this might change again in 2015.

to a large extent to language. Type 2 concepts, as we take them, are explicit ways of making categories, and this often involves language, i.e., making distinctions using language. However, what we *do not* want to argue is that *all and only* processes involving Type 2 concepts are linked to language. On the one hand, some Type 1 concepts are linked to language, especially when some explicit distinctions become overlearned or, more prosaically, when we speak daily—we do not need to pick out every single word we say. On the other hand, Type 2 concepts, as we said, and will continue to argue in the rest of this section, are involved in automatic (albeit probably slow) Type 1 processes, shaping the content of Type 1 concepts, especially in learning.

As we agreed above, definitions may not be central in the fast pace of day to day cognition. For the reasons expressed above, however, we would certainly not characterize them as "marginal" as Machery (2010, 2011) does. Definitions, he argues, are only used intentionally in particular conditions. Moreover, he believes there has to be cognitive control exercised upon tasks in this condition to inhibit the more automatic processes (e.g., using prototypes, exemplars and theories). For Machery, we rarely use definitions, even in cases where there are clear and formal definitions available:

> I suspect people rarely use these definitions in reasoning, to categorize, and so on. Instead, people seem to use prototypes, exemplars, or causal theories of grandmothers or bachelors. Evidence is consistent with these suspicions. (Machery, 2011, p. 207)

This is surely true, but not an objection to our proposed view. Definitions sometimes are used in cognition and they sometimes influence the acquisition of Type 1 concepts—especially when prototypes and exemplars can mislead, as is the case for many concepts in science and other formal contexts (judiciary, business, technological)—and we do not claim more than this.

In the context of what we have presented, to eliminate Type 2 concepts from concept science would amount to excluding many important processes that are causally relevant to the processes studied in this domain. The argument that Type 2 concepts should be excluded from the psychology of concepts because they are not "used by default" seems ad hoc once the natural kind assumption and single-process views of the mind are rejected. We think there is a case to be made for the inclusion of other kinds of concepts, such as those we started to describe here. If concepts have both a Type 1 and a Type 2 aspect and if, as we argued, they interact causally, then both must be studied.

## 4.6 Methodological challenges for Type-2 concepts

We have seen one reason why definitions as they are used in the psychological projects were thought to be necessary and sufficient conditions: the conflation with the scientific project for definitions. Now we will see why *any* kind of definition was bound to be inadequate in the psychology of concepts.

Definitions are at an explanatory disadvantage in a framework marked by the natural kind assumption and a single-process view of the mind. It is noteworthy that an important objection to the classical theory of concepts is that definitions do not account for typicality effects. Typical members of a set are categorized more quickly than atypical members. Central to these experiments, then, is participants' response-time. Under the view that only one type of body of information can play the role of concepts (the natural kind assumption), and the view that experiments emphasizing reaction time are relevant to understand any cognitive process (the single-process view), such results show that definitions cannot be concepts. By devising experiments where participants must respond very quickly, current experimental standards look at very small time differences between actions within a task that participants must accomplish. This allows automatic processes to shine, and more importantly be studied, but this does not allow for a complete understanding of the bodies of information used by human beings to categorize. Definitions are *not used by default* by cognitive processes—partly because they are relatively slow, but also because they don't use the same underlying processes. These special properties must be taken into account if we want to study definitions as bodies of information used in cognitive processing. If we are right in claiming that definitions should be considered as an interesting type of body of information in concept science, we will have to change the working assumptions of this science. The defining characteristics of concepts, in this new framework, will not only be processing speed, but also precision and accuracy. So-called gut feelings retain their importance, but reflective processes would be studied for what they do best.

Definitions are explicit ways of making categories. Sometimes, as in science, we may wish to construct necessary and sufficient conditions; other times, as when we correct our child's usage, one easily observable criterion such as the dolphin's playfulness may suffice. The current literature on concepts adopts a view that makes it nearly impossible to study definitions as psychological processes. Important changes in concept science will have to be made if we are to learn more about definitions. As a first sketch, however, we can say that definitions are to be contrasted with automatic processes such as those behind prototypes, exemplars and theories. Definitions are those concepts we that learn explicitly (e.g., by interacting with others, reading) and that we apply in a reflective fashion.

In the two following sections, we first explain how experiments in concept science are designed to favor Type 1 processes, before addressing a more specific worry about how developmental psychologists construct different types of theories, using their "theory" theory approach of categorization.

### 4.6.1 Experiments in concept science

We mentioned in passing that definitions are at a disadvantage in current experimental frameworks in cognitive science. The reason is simple: if there is one, and only one, categorization process at work, the best results should be obtained by having participants do the task as quickly and with as little reflection as possible, so as to avoid involvement of other cognitive processes ("isolating the process"). The task should also be as new and abstract as possible, i.e., far away from participants' every day experiences, using stimuli that they will not have encountered before. This second requirement ("avoid contamination") is in place in order to avoid variation between participants: using well-known animals, for example, could change the results for participants who have had extensive contact with the species in the experiment[45]. While uncovering such concept formation processes is important and interesting, there are many ways in which we come to form concepts, as we argued previously.

An example will illustrate these requirements. In defending the prototype model, Minda & Smith (2001) suggested that previous results from experiments supporting the exemplar model were due to the type of stimuli presented. These stimuli were

45. This is a worry we should keep in mind while designing new experiments and interpreting their results when working on definitions. Variation will very likely be great between participants depending, among other things, on their educational and cultural backgrounds.

varied along four or fewer dimensions (see Figure 1). Minda & Smith's contention is that, for more complex stimuli—they were using bug-like creatures, the exemplar model was not performing as well as their prototype model[46]. For this reason, they presented their participants with six- and eight-dimensional stimuli (e.g., Figure 2).

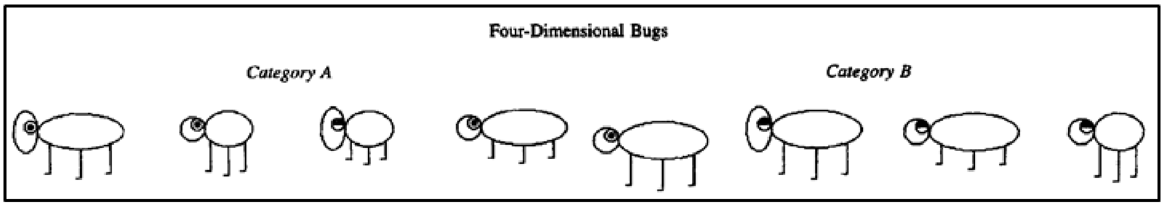


Figure 1. Four-dimensional stimuli used in Minda & Smith (2001, p. 797)
© 2001 American Psychological Association, Inc.

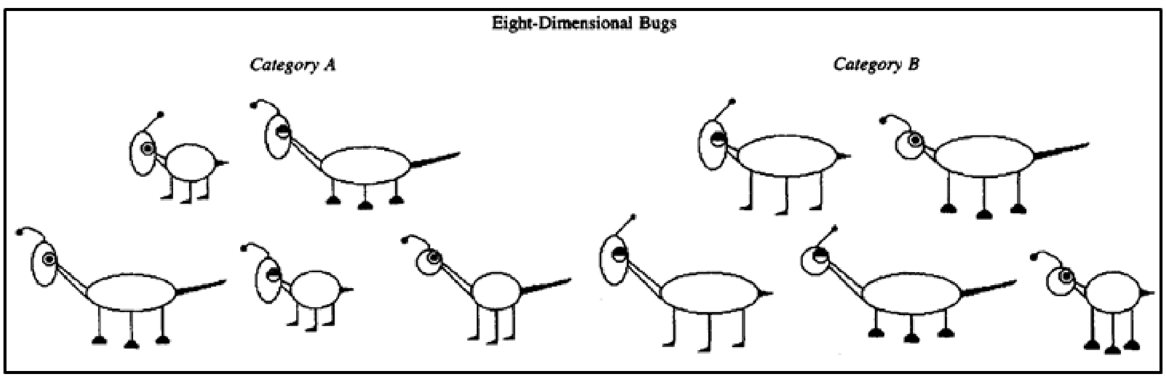


Figure 2. Eight-dimensional stimuli used in Minda & Smith (2001, p. 797)
© 2001 American Psychological Association, Inc.

These "bugs" were designed with a third requirement in mind: it has to be hard to formulate an explicit rule to differentiate Category A from Category B while completing the task (see Minda, Desroches & Church's (2008) discussion of what happens when participants start formulating explicit rules; we will say more on this shortly). As the stimuli in Figure 1 and in Figure 2 show, the categories do not vary along only one dimension. Long legs will not help differentiate between A and

46. "research on exemplar theory has often featured small, poorly structured categories containing low-dimensional stimuli" (Minda & Smith, 2001, p. 775)

B, and the same feature for eyes will be found in both categories A and B. This is a feature implemented *specifically* to isolate prototype- or exemplar-like features—an important feature of the experiment in a context where they were trying to isolate a given type of categorization process. Minda & Smith relied here on well-known limitations of memory and attention. It is very hard to keep in mind more than four or five elements at once (the "magic number" is indeed smaller than initially supposed by Miller (1956), cf. Cowan (2010)), something that would be necessary to formulate rules to correctly classify Minda & Smith's creatures.

In the trials, participants were presented with a drawing of a bug and had to determine to which category it belonged, a choice followed by a sound signaling success or failure. Although participants did not have time constraints to complete the task, the stimuli were complicated enough that it was indeed very hard to derive a rule from them (and the participants were neither encouraged nor discouraged to do so). Moreover, the participants had to complete, in Minda & Smith's (2001) Experiment 1, 560 trials, and 960 trials for Experiments 2 to 4 (p. 778), making it unlikely that participants would spend a long time pondering over each trial.

This kind of experiment favors prototypes for various reasons: while the exemplar theory predicts that the process computes overall similarity and demands memorization of elements of the set and further comparisons of how they look (which works well, e.g., when dimensionality is low, or when stimuli are similar enough to one another), the task designed by Minda & Smith (2001) makes this very hard by having higher dimensionality. The same is true for making explicit rules on the fly: higher dimensionality makes it unlikely that participants will be able to formulate useful categorization rules (if it is, at all, possible). Prototype theory rather predicts that the categorization process functions by statistically monitoring various features and aggregating automatically those in two groups. This is what Minda & Smith's (2001) results show.

When participants are given a chance to formulate rules to help them with their categorization, we see interesting patterns arising. Minda and colleagues (2008) conducted experiments with both children and adult participants in order to see how this would affect their overall categorization processes. Using four of Shepard, Hovland, & Jenkins' (1961) six stimuli sets, they compared how children (3-, 5-, and 8-years-old) and adults categorize. An important difference between these experiments and the ones discussed above, first, is that the Shepard and colleagues' (1961) sets vary only according to three dimensions (shape, color, size)—making them much easier to track than the six or eight dimensional bug creatures used in Minda & Smith (2001). In order to make our argument, we now have to describe these different ways of making the distinctions between two categories.

Minda and colleagues (2008) used single dimension (SD), disjunctive (DR), nonlinearly separable (NLS) and family resemblance (FR) sets. The SD set varies along only one dimension and is easy to formulate with a rule (i.e., if feature A, then Category 1), but is also easy to grasp by other categorization processes (i.e., exemplar- or prototype-based, since they are specialized in picking up similarities and overall resemblance, see Figure 3).

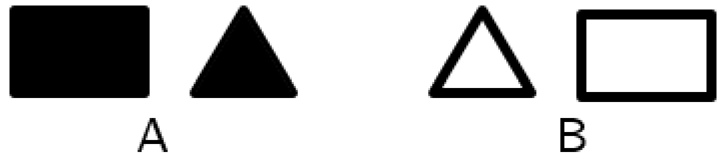


Figure 3. Example of a single dimension (SD) distinction

The DR set, in a classic version of the task, puts black triangles and white squares, no matter their size, in one category, and the white triangles and black squares in the other (see Figure 4). Even if the participant has to keep in mind only two dimensions in the DR case, it is hard to come up with the correct rule (and Type 1 processes, specifically exemplars and prototypes in this case, are not of much help because it is not possible to rely on any type of resemblance) to learn the two categories because the structure is misleading; when the participant finds what the rule is, however, the categorization task becomes easier.


Figure 4. Example of a disjunctive rule (DR) distinction

NLS refers to a set that implements a rule with an exception. The example used by Minda and colleagues (2008, p. 1520) is “black objects and the small white triangle” for Category 1 (see Figure 5). NLR is hard because applying the rule demands considerable cognitive resources to take into account the exception, including attending to all three dimensions of the objects at all times. The nonlinearity also makes it very difficult to learn without any appeal to explicit rules, i.e., by using only Type 1 processes.

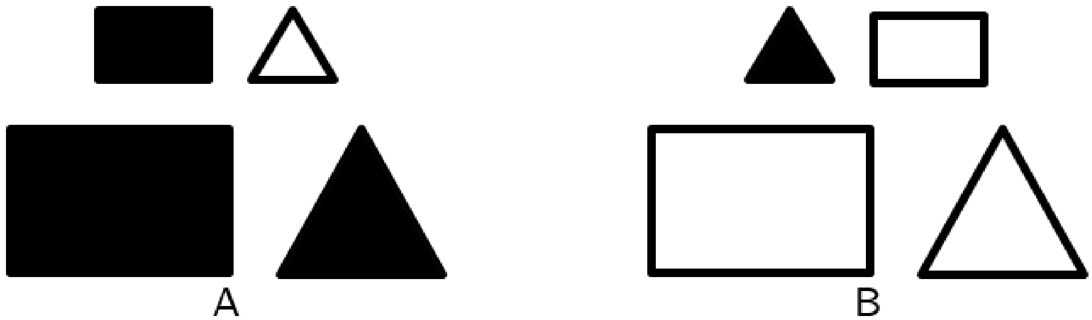


Figure 5. Example of a nonlinearly separable (NLR) distinction

Lastly, similar to NLR at first glance, the FR set is designed so that members of a category share a majority of their features with the other members of that category. As Minda and colleagues point out, such sets can also be described by complex rules of the form "share any two of the following three features", but the fact that the members of a category *look alike* makes it possible for Type 1 processes to pick up the distinction—relying either on general resemblance, if prototype and exemplar processes are picking it up, or relying on statistical regularities to form Bayesian nets in the case of theories (see Figure 6). This latter characteristic is the characteristic that the researchers used to design the six and eight dimensional bugs. To sum up, Type 1 processes as we described them above can easily latch on to features of each category in the SD and FR cases, and rules seem to be needed to perform well in the DR and NLS cases, a claim that the results detailed in Minda and colleagues' (2008) paper seem to support.

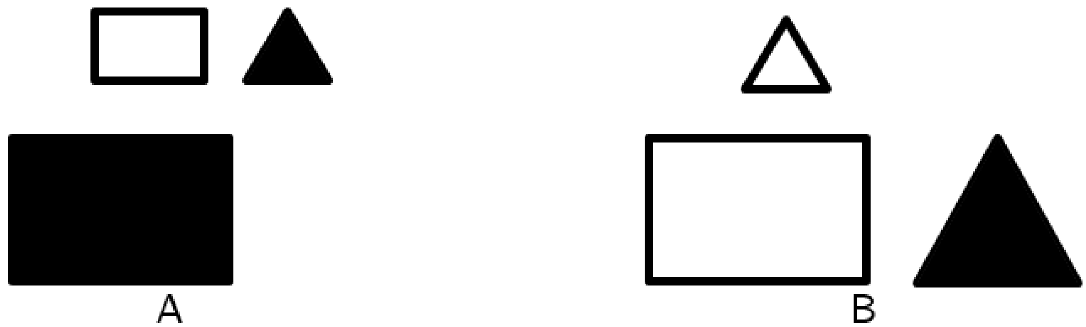


Figure 6. Example of a family resemblance (FR) distinction

The results of their Experiment 1 show a tendency where there is a marked difference between children and adults when it comes to formulate and to apply the rules in the DR and NLS cases—the proportion of correct answers given by children drops below 70% for DR, and is slightly above 75% in NLR. The adult participants give >90% correct answers in both these cases, and show a clear learning curve (at some point, they "get it"). Minda and colleagues (2008, p. 1524) summarized their results as follows: "children generally lagged behind adults when learning categories that depended on complicated verbal rules but not when learning categories that required a simple rule, or when the categories did not depend on verbal rules". These results are in line with Ashby and colleagues' (1998), where adults seem to default to using verbal rules in most learning conditions, and when they do not do so, as in the FR case, we find children's and adults' performances to be at the same level. This also suggests that mastering language can be key in processing certain kinds of categorization tasks—children between 3 and 8 definitely use language, but ease of use helps as it leaves more cognitive resources to other critical skills such as inhibiting automatic responses in favor of more controlled ones, which is one factor that explains differences between children's and adults' performances in DR and NLR. For Minda and colleagues, their Experiment 1 "provided the first developmental evidence for multiple category learning systems" (Minda et al., 2008, p. 1530).

What we find especially interesting in such experiments, for our current purposes, is that processes along the lines of what we call definitions can shine in the proper context. The developmental differences are also of interest as adults (viz., undergraduate students at the University of Western Ontario, in the above case) are used to employing such categorization processes in their everyday lives. Our contention against Minda and colleagues' (2008) analysis[47], however, is that the distinction is made along the verbal / nonverbal axis, something we do not think will hold, as the interactions are much more complex than what this suggests (see also Piccinini (2011) and Machery's (2011, sec. 7, 9) response). It is not as if the verbal processes were not using what we identify as Type 1 concepts, or that all verbal processes are of the type we have in mind when we talk about definitions, viz. we think that there are processes that are verbal but do not have to do with definitions. There is, here, much more ground to explore: we need to understand how these different types of processes interact but only by keeping in mind the difference between Type 1 and Type 2 concepts will we be able to come up with a good general model of category formation, category learning and, ultimately, a better, more complete, account of concepts such as the one we are sketching here.

Before bringing this article to a close, there is one last issue we want to discuss, one raised in Machery's (Machery & Seppälä, 2009; Machery, 2011) response to remarks similar to the ones we make here (e.g., Poirier & Beaulac, 2011). There

47. This criticism also applies to Ashby et al.'s (1998) COVIS ("competition between verbal and implicit systems") model (see also Ashby & Maddox (2005)).

are contentions regarding what “theories” are within the theory theory approach in developmental psychology, and we want to suggest that the best available versions of the theory theory do not allow us to explain the type of phenomena we have put forward in this paper. The theory theory, we will argue, squares well with our characterization of Type 1 concepts, and does not allow proper explanation of what is going on under the umbrella of what we call Type 2 concepts.

### 4.6.2 The theory theory approach and definitions

A difficulty we encounter within the theory theory paradigm is that it remains rather vague what a “theory” is supposed to be in the first place. Some analyses of this way of looking at (especially children’s) categorization processes seemed to be pretty clear about the idea that we should think about theories as being along the model of a scientist’s formulation of a theory (Keil, 1989b), although not an explicit one—viz. children do not formalize or verbalize these theories but act as if they were holding such theories, and forming, testing and confirming hypotheses (Gopnik, Meltzoff, & Kuhl, 2001). Most recent accounts, however, seem to think in terms of Bayesian nets, causal maps developed automatically by systems in charge of spotting regularities and correlations in the children’s environment (Gopnik et al., 2004), thus offering a more formal and an easier to test model of what concepts are in this paradigm. This latter approach to theory is also in line with what we have been saying about Type 1 concepts in the previous sections of this paper: on this view the theory theory approach would describe one other aspect of (or a different kind of) Type 1 concepts.

We will take these two questions in turn: what a theory is and whether we can think of theories as being Type 1 concepts. We will then turn to Machery’s idea that we do not need a new body of information (other than prototypes, exemplars and theories) in order to explain cases such as those we outlined above because theories could, according to him, explain the phenomena we have been interested in (Machery & Seppälä, 2009; Machery, 2011).

A first brush at what a theory is in this context should start by pointing out what this way of thinking about categorization is supposed to explain: prototype- and exemplar-based accounts of categorization rely mostly on similarity relationships between objects, and there are features of categorization processes that seem to remain unexplained under such a view. The theory approach brings to the discussion other types of relations we use in order to group objects and creatures. This is why another animal cleverly disguised as a skunk *is not* mistaken to be a skunk, even according to very young children (Keil, 1989a). This supports the idea that children seem to be forming assumptions about what makes it the case that a given animal is, or is not, a skunk. This challenge to the prototype- and exemplar-based approaches is one of the main motivations behind the theory theory: why isn't it the case that superficial resemblances are enough to categorize a non-skunk as a skunk? There are multiple other advantages to these models, such as their being good at explaining why we are sometimes able to automatically and easily form categories that do not seem to be linearly separable, why we expect to observe similar behavior or features in objects we put in the same category even if they are not superficially similar, and why the causal connections between features are picked up so fast and can be transferred from one category to the other without any apparent difficulty. This last point is, in fact, the main point put forward by proponents of the theory theory; they argue that children form causal maps that are Bayesian in nature[48].

This model, probably the most plausible one currently available to describe what a theory is, is developed by Gopnik and colleagues (2004). Of course, as they say, adults can rely on "substantive prior knowledge" (Gopnik et al., 2004, p. 7) when they make physical and psychological predictions. "Formal assumptions" can be used—explicitly, we assume, given the context where it is used in the paper—as well in order to learn new correlational patterns. This is not, however, the kind of phenomena that the causal maps account tries to explain[49]. Rather, in contrast, they are trying to cash out what the "formal assumptions" that children might be using *implicitly* are—hence why we take the first use of this expression to mean that the formal assumptions are made explicit in the adult case. These causal relations are represented, they argue, along lines described by Bayes' net formalism. The advantage of such a framework is that it allows a better explanation of how

48. Churchland (1989) suggests a similar analysis of theories against "classical views".

49. Whether children can or cannot use "formal assumptions to recover causal maps from patterns of correlation between events" (Gopnik et al., 2004, p. 7) remains an open question for Gopnik and colleagues. We assume, given our theoretical framework, that they might do so to a certain extent, and starting at a certain age—as we saw with Minda and colleagues (2008), this is a hard task to do even for an 8 year old. There is still a need to distinguish this kind of formal learning from the learning of causal maps, which relies on some innate mechanisms in addition to "substantive knowledge" acquired gradually on Gopnik and colleagues' view. Indeed, they specify that learning some knowledge-that,

> that remote controls activate television sets, that watering plants makes them grow, or that crowds make shy people nervous", could "play a major role in the impressive changes in causal knowledge we see in the development of everyday theories. (Gopnik et al., 2004, p. 7)

We believe this prediction to be exactly right, and our goal is to offer a framework that could allow for such explanations in developmental psychology down the road, and to do so by considering carefully the kind of role structures such as those we identify here can have.

two possible causes of a given phenomenon, or causes given certain circumstances, are weighted against one another when planning a course of action.

The overall concern here is that "theory" within the theory theory approach has an ambiguous meaning; it can both mean a theory of a scientific kind or a Bayesian causal map formed automatically. We do not aim at establishing which the correct one is, but we suspect that a natural kind assumption might be at play here as well. There might very well be more than one body of information called "theory" in the literature as it currently stands.

However, according to some of the more recent and precise views on the matter—of which Gopnik and colleagues (2004) are representative (see also the papers in Gopnik & Schulz, 2007)—theories are Bayesian causal nets, constructed automatically by cognitive processes that track recurrence of events and how they relate to one another. This framework's goal is to give an account of how the causal structure of the world is represented by children. Although there are other accounts of "theory" in the literature, this is the one we have in mind when we associate the theory theory approach to Type 1 concepts. Importantly, since for Machery (2009) concepts are bodies of information used by default, this is also the kind of body of information he would have to refer to when he compares the theory theory approach to the prototype and exemplar based proposals.

Yet, Machery insists. In his paper with Seppälä (Machery & Seppälä, 2009), he claims that the role we assign here to definitions—explicitly made assumptions about objects, but also word-association and scientific theorizing on some concepts—is to be filled by the theory theory approach in his heterogeneity hypothesis (Machery, 2011). This is how they explain that "tomato" is both a vegetable and a fruit, “whale” both a mammal and a fish, and “zombies” are both dead and alive—presented as conflicts between similarity-based bodies of information (prototypes and exemplars) and theories, or possibly even between two theories. This ambiguity in what “theory” means in this context helps Machery make his argument. One can’t have this both way. Yes, “theory” can refer, within the theory theory paradigm, both to automatically constructed Bayesian nets built by processes that compute co-occurrence and causal relationships in the world (Gopnik et al., 2004), and to a body of information similar to scientific theories with explicitly held beliefs about the structure of the world and what makes some objects members of a category—a skunk is not a skunk only in virtue of being a black animal with a white stripe, it has to have certain characteristics, a *skunkeness*, that is related to the kind of animal it is; what is inside counts, not only the appearance (Keil, 1989a). We do not claim each of these kinds of “theory” do not have effects on one another. The way we see Type 2 concepts makes it possible, if not mandatory, that there would be influences of one on the other, and vice-versa.

What we want to conclude from this discussion is that Machery’s assumption that theories can explain the type of phenomena he identifies in Machery & Seppälä (2009) or that we discuss here (see also Poirier & Beaulac, 2011) cannot be right. He uses the ambiguity of the notion of “theory” to avoid acknowledging that some of the most influential theory theory accounts are about processes *used by default*, something motivated by experimental biases illustrated by the example discussed in section 4.6.1.

Our claim in this paper has been that the bodies of information that are used by default do not and cannot explain fully human categorization. In this sense, the theory theory approach can be understood under the umbrella of the other bodies of information discussed previously and account for a very important part, probably even the main part, of how human beings categorize. Yet, it might miss other kinds of bodies of information that could have a role, such as what we identified here as Type 2 concepts—to refer mostly to what we call definitions.

## 4.7 Yes, but why "definitions"?

Someone could accept everything we have said above but insist: But why do you insist on calling Type 2 concepts (or one part of what Type 2 concepts are) "definitions". The short answer is because this is what they are, and because the dynamics of science allows for conceptual change. The fact that a group of scientists mistakenly used a term for a period of time has never been (in point of descriptive philosophy of science) and should not be (in point of normative philosophy of science) sufficient for banning forever the use of the term.

Moreover, the word "definition" has a widespread use that refers to the kind of things that explain the meaning of words in a dictionary. We think this description of a definition sits well with the account we have defended here. It is not defined in necessary and sufficient terms or as a formal entity. Yet, the structure of these categories might reveal more than we first thought about the organization of our minds and on how we navigate the world. Definitions, as they are in a dictionary, are definitely interesting, but we understood the term even more broadly in 4.4: they can serve as a guide to better understand the world, but also as a tool to form ad hoc categories for different purposes ("objects on the left" vs "objects on the right" without any dependence on similarity or other features than the relative position of the objects to those having a discussion). Some researchers have even started to look into the relationships between words within dictionaries to see what it tells us about the structure of our languages (Blondin-Massé et al., 2008; Blondin-Massé, Harnad, Picard, & St-Louis, 2013). A very interesting route for future research, alongside those we mentioned in 4.3.1, would be to investigate how (and if) this translates into individual speakers' conceptual structures—using "definitions" for Type 2 concepts makes it possible to link these perspectives on conceptual structure seamlessly.

Our view also opens up the possibility that some of the outstanding issues in current debates concerning the notion "concept" result from a conflation of different roles that concepts are asked to play, especially if Type 2 roles are wrongly attributed to Type 1 concepts and vice versa (see, e.g, our discussion in section 4.5). The rehabilitation of definitions in concept science also points towards a gap in our understanding of concepts: the interface between language and concepts is rarely studied or, if it is, it seems that the important distinctions we suggested in the present paper are not made. Finally, rehabilitating definitions may help close the gap Machery (2009) opened between the psychological and the philosophical use of the notion of concept. Schneider (2010), for example, suggests that Conceptual Atomism can accommodate both perspectives on concepts; Rey (2010) rather thinks that externalism offers such a framework where prototypes, exemplars and theories are the means by which one arrives at a concept (in the philosophical sense). We do not want to judge the value of these claims here, but we think the kind of distinction we bring forward can help this discussion by making clear what the roles of different psychological processes might be, and how they fit within a broader view of concepts.

## 4.8 Conclusion

In this paper, we argued that rejection of the natural kind assumption and adoption of a dual-process view of the mind allow for the rehabilitation of definitions—understood broadly as one of the Type 2 bodies of information psychologists must study in order to flesh out a full account of concepts. If we are right, this should have methodological consequences on some of the work done in concept science, especially in opening-up new areas of research favoring novel experimental designs; it is interesting (and pleasing to us) to observe that some psychologists are developing frameworks compatible with our view (Minda & Miles, 2010). Some of this work is even showing that—as it could be predicted from the perspective of recent accounts of dual-process theories where working memory is thought to have a central role in Type 2 processes (Evans & Stanovich, 2013)—there is a correlation between rule-based categorization, the execution of Type 2 processes, and working memory capacity (Miles, Matsuki, & Minda, 2014; Rabi & Minda, 2014).

## References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*(3), 442–481.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56*, 149–178. doi:10.1146/annurev.psych.56.091103.070217

Bird, A., & Tobin, E. (2012). Natural Kinds. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2012.). Retrieved from http://plato.stanford.edu/archives/win2012/entries/natural-kinds/

Blondin-Massé, A., Chicoisne, G., Gargouri, Y., Harnad, S., Picard, O., & Marcotte, O. (2008). How is meaning grounded in dictionary definitions? In *Proceedings of the 3rd Textgraphs Workshop on Graph-Based Algorithms for Natural Language Processing* (pp. 17–24). Association for Computational Linguistics.

Blondin-Massé, A., Harnad, S., Picard, O., & St-Louis, B. (2013). Symbol Grounding and the Origin of Language: From Show to Tell. In C. Lefebvre, H. Cohen, & B. Comrie (Eds.), *Current Perspective on the Origins of Language*. Benjamin. Retrieved from http://eprints.soton.ac.uk/271438/

Borges, J. L. (1972). *El hacedor*. Madrid: Alianza Editorial.

Brooks, L. R., & Hannah, S. D. (2006). Instantiated features and the use of "rules." *Journal of Experimental Psychology. General, 135*(2), 133–151. doi:10.1037/0096-3445.135.2.133

Brooks, L. R., LeBlanc, V. R., & Norman, G. R. (2000). On the difficulty of noticing obvious features in patient appearance. *Psychological Science, 11*(2), 112–117.

Brooks, L. R., Squire-Graydon, R., & Wood, T. J. (2007). Diversion of attention in everyday concept learning: Identification in the service of use. *Memory & Cognition, 35*(1), 1–14. doi:10.3758/BF03195937

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Churchland, P. M. (1989). On the Nature of Theories: A Neurocomputational Perspective. In W. Savage (Ed.), *Scientific Theories: Minnesota Studies in the Philosophy of Science* (Vol. 14, pp. 59–101). Minneapolis: University of Minnesota Press. Retrieved from http://www.mcps.umn.edu/philosophy/14_4churchland.pdf

Cowan, N. (2010). The Magical Mystery Four: How Is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science, 19*(1), 51–57. doi:10.1177/0963721409359277

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology, 59*(1), 255–278. doi:10.1146/annurev.psych.59.103006.093629

Evans, J. S. B. T., & Frankish, K. (Eds.). (2009). *In Two minds : Dual processes and beyond*. New York: Oxford University Press.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science, 8*(3), 223–241. doi:10.1177/1745691612460685

Fodor, J. A. (1983). *The modularity of mind : an essay on faculty psychology.* Cambridge (MA): MIT Press.

Fodor, J. A. (2000). *The mind doesn't work that way: the scope and limits of computational psychology.* Cambridge (MA): MIT Press.

Fodor, J. A., Garrett, M. F., Walker, E. C. T., & Parkes, C. H. (1980). Against definitions. *Cognition, 8*(3), 263–367. doi:10.1016/0010-0277(80)90008-6

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673–1682.

Gawronski, B., & Creighton, L. A. (2013). Dual-Process Theories. In *The Oxford handbook of social cognition.* New York: Oxford University Press. Retrieved from http://publish.uwo.ca/~bgawrons/documents/GC_Oxford.pdf

Gawronski, B., Sherman, J. W., & Trope, Y. (2014). Two of What? A Conceptual Analysis of Dual-Process Theories. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 3–19). New York: Guilford Press.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*(3), 183–209. doi:10.1016/0010-0277(86)90034-X

Giere, R. N. (2006). *Scientific perspectivism.* Chicago: University of Chicago Press.

Gigerenzer, G., & Regier, T. (1996). How do we tell an association from a rule? Comment on Sloman (1996). *Psychological Bulletin, 119*(1), 23–26. doi:10.1037/0033-2909.119.1.23

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological Review, 111*(1), 3–32. doi:10.1037/0033-295X.111.1.3

Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. (2001*). The scientist in the crib : what early learning tells us about the mind.* New York: HarperPerennial.

Gopnik, A., & Schulz, L. (Eds.). (2007). *Causal learning. Psychology, philosophy, and computation.* New York: Oxford University Press.

Griffiths, P. E. (2004). Emotions as Natural and Normative Kinds. *Philosophy of Science, 71*(5), 901–911.

Kahneman, D. (2011). *Thinking, fast and slow.* Toronto: Doubleday Canada.

Katz, J. J., & Fodor, J. A. (1963). The Structure of a Semantic Theory. *Language, 39*(2), 170. doi:10.2307/411200

Keil, F. C. (1989a). *Concepts, kinds, and cognitive development.* Cambridge (MA): MIT Press.

Keil, F. C. (1989b). Spiders in the Web of Belief: The Tangled Relations Between Concepts and Theories. *Mind & Language, 4*(1-2), 43–50. doi:10.1111/j.1468-0017.1989.tb00239.x

Keren, G., & Schul, Y. (2009). Two Is Not Always Better Than One: A Critical Evaluation of Two-System Theories. *Perspectives on Psychological Science, 4*(6), 533–550. doi:10.1111/j.1745-6924.2009.01164.x

Kim, J. (1992). Multiple Realization and the Metaphysics of Reduction. *Philosophy and Phenomenological Research, 52*(1), 1. doi:10.2307/2107741

Ladyman, J., & Ross, D. (2007). *Every thing must go: metaphysics naturalized.* New York: Oxford University Press.

Machery, E. (2005). Concepts Are Not a Natural Kind. *Philosophy of Science, 72*(3), 444–467.

Machery, E. (2009). *Doing without Concepts.* Oxford; New York: Oxford University Press.

Machery, E. (2010). Replies to my critics. *Philosophical Studies, 149*(3), 429–436. doi:10.1007/s11098-010-9530-3

Machery, E. (2011). Replies to Lombrozo, Piccinini, and Poirier and Beaulac. *Dialogue, 50*(1), 195–212. doi:10.1017/S0012217311000199

Machery, E., & Seppälä, S. (2009). Against Hybrid Theories of Concepts. *Anthropology & Philosophy, 10*(1-2), 99–127.

McClelland, J. L., & Rogers, T. T. (2004). *Semantic cognition: a parallel distributed processing approach.* Cambridge (MA): MIT Press.

Miles, S. J., Matsuki, K., & Minda, J. P. (2014). Continuous executive function disruption interferes with application of an information integration categorization strategy. *Attention, Perception, & Psychophysics*, *76*(5), 1318–1334. doi:10.3758/s13414-014-0657-8

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97. doi:10.1037/h0043158

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and non-rule-described categories: a comparison of children and adults. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 34*(6), 1518–1533. doi:10.1037/a0013355

Minda, J. P., & Miles, S. J. (2010). The Influence of Verbal and Nonverbal Processing on Category Learning. In B. H. Ross (Ed.*), Psychology of Learning and Motivation* (Vol. 52, pp. 117–162). Elsevier. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0079742110520036

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(3), 775–799. doi:10.1037//0278-7393.27.3.775

Piccinini, G. (2011). Two Kinds of Concept: Implicit and Explicit. *Dialogue, 50*(1), 179–193. doi:10.1017/S0012217311000187

Piccinini, G., & Scott, S. (2006). Splitting Concepts. *Philosophy of Science, 73*(4), 390–409.

Piccinini, G., & Virtel, J. (2010). Are Prototypes and Exemplars Used in Distinct Cognitive Processes? *Behavioral and Brain Sciences, 33*(2-3), 226–227.

Pitt, D. (1999). In defense of definitions. *Philosophical Psychology, 12*(2), 139–156. doi:10.1080/095150899105846

Poirier, P., & Beaulac, G. (2011). Le véritable retour des définitions. *Dialogue, 50*(1), 153–164. doi:10.1017/S0012217311000163

Prinz, J. J. (2004). *Furnishing the mind : concepts and their perceptual basis.* Cambridge (MA): MIT Press.

Rabi, R., & Minda, J. P. (2014). Rule-Based Category Learning in Children: The Role of Age and Executive Functioning. *PLoS ONE*, *9*(1), e85316. doi:10.1371/journal.pone.0085316

Rey, G. (2010). Concepts versus conceptions (again). *Behavioral and Brain Sciences, 33*(2-3), 221–222. doi:10.1017/S0140525X10000440

Samuels, R. (2009). The magical number two, plus or minus: Dual-process theory as a theory of cognitive kinds. In J. S. B. T. Evans & K. Frankish (Eds.), *In Two Minds. Dual Processes and Beyond* (pp. 129–146). New York: Oxford University Press.

Schneider, S. (2010). Conceptual atomism rethought. *Behavioral and Brain Sciences, 33*(2-3), 224–225. doi:10.1017/S0140525X10000579

Schwartz, S. P. (1979). Natural kind terms. *Cognition, 7*(3), 301–315. doi:10.1016/0010-0277(79)90003-9

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied, 75*(13), 1–42. doi:10.1037/h0093825

Stanovich, K. E. (2011). *Rationality and the reflective mind*. New York: Oxford University Press.

Toates, F. (2006). A model of the hierarchy of behaviour, cognition, and consciousness. *Consciousness and Cognition, 15*(1), 75–118. doi:10.1016/j.concog.2005.04.008

Van Gelder, T. (1995). What Might Cognition Be, If Not Computation? *The Journal of Philosophy, 92*(7), 345. doi:10.2307/2941061

Weiskopf, D. A. (2009). The Plurality of Concepts. *Synthese, 169*(1), 145–173.

Yarbus, A. L. (1967). *Eye Movements and Vision*. New York: Plenum Press.

# Chapter 5 : Conclusion

*Most of the fundamental ideas of science are essentially simple, and may, as a rule, be expressed in a language comprehensible to everyone.*

Albert Einstein & Leopold Infeld (1938)

# 5 Conclusion

Churchland (2012, p. 252) writes that it is ironic that theories, such as his, that are "*pre*linguistic/*sub*linguistic" accounts of cognition make it possible to better appreciate how language can be transformative for cognition. I think this remark is justified insofar as the assumption that language has no role, or that it is so closely tied to thoughts that distinguishing between the two would be a waste of time, is commonplace in many philosophical perspectives. Arguing for the validity of this thesis is Churchland's work, and I have not attempted to do this here, but I believe I have offered a glimpse into the kind of consequences this lack of attention to language brings about. In other words, if we think that, at its very core, cognition is language-like, we will be ignoring important transformations made possible by language—this is a lesson I take from my discussion of dual-process theories and how the question of language is sometimes put aside in this work. I hope I have argued convincingly in this dissertation that researchers often attribute various roles to language without looking closely enough at the underlying assumptions at play, and that making these assumptions clearer can help research move forward. Likewise, I argued for the hypothesis that language might have many roles at once in cognition, roles that might differ depending on the cognitive process that language interacts with, and depending on which part(s) of the faculty of language are doing this interaction.

There is still much to be understood about the language faculty, how it is divided, and how it came about, but I think it is clear it played a major role in how human cognition changed, from an evolutionary point of view, from that of other species. The main take-home message of the work presented here should be that we will not be able to answer this question by proposing a single role for language, a position that goes against a trend in many previous attempts in philosophy and in psychology—but also one that seems prevalent in the literature on the evolution of language. If this is right, there are no good answers to the question "What is the role of language in cognition?" and attempting to find an answer will not help move research forward. Accepting this idea would help avoid very unproductive, unfruitful, debates about roles language might have.

The first paper was a sketch of the landscape of these various roles for language, and how this sketch differs from two earlier attempts by Carruthers (2012; Botterill & Carruthers, 1999). In addition, I argued that we should not confine ourselves to a single role for language and that we should not see these various accounts as being part of a single continuum of positions, from one extreme to the other. Language is not just a tool having more or less important effects on cognitive processes; it permeates cognition and changes how we interact with the world. Confining ourselves to the view that language has a single role, as I argued, would have (and has had) many adverse effects. Although my account is still imperfect, and does not fully cover this immense literature, I believe the account I presented can, at least, be used as a basis for future discussions. I hope it will at least be useful to offer guidelines (if only as an example of what not to do!) for further refinements—I will pursue this work and I hope other researchers will join me in investigating these complex questions.

In the second paper, I discussed assumptions about language made in some of the most influential frameworks in the architecture of mind literature. Many dual-process theorists are not careful enough when they consider the role of language in the architecture, and I suggested that dividing language into its many components might help us gain some traction on this problem. The goal of this paper, within the context of this dissertation, was to show that the concerns I raised while discussing (mainly) philosophical accounts of the relations between language and cognition were also present in work in cognitive science—namely, in this case, in psychology. In the third paper, I then discussed a specific case to illustrate how this kind of reasoning about cognitive processes could be applied to the investigation of one cognitive capacity: categorization processes.

I chose categorization processes for two reasons. First, it is *one* area of research in psychology and, although it might be the investigation of many underlying, independent processes (Machery, 2009), it still did not force me to commit to roles language might have in other cognitive processes—I did not have to adopt a *generalized* account of the architecture of mind (following the terminological suggestion made by Gawronski & Creighton, 2013). In this sense, I did not need a theory of the role of language in cognition in general, but only a theory of the role of language *in categorization processes*. Second, although I assumed language was one process in the context of this third paper—against a suggestion of my own made in the second paper[50], language does not have a role constrained to one type of process in my proposed account of definitions. It cuts across both Type 1 and Type 2 processes, illustrating the complex relationship dual-process theories have with language. Hopefully, the account I have developed with Pierre Poirier helps us to understand how each kind of process does its work in the context of categorization.

50. Although this is a problem, this choice was motivated in the present context by two concerns: the length of the paper and the (un)availability of empirical data that would allow a careful discussion of how various parts of FLB interact with categorization processes.

Accordingly, in the case of categorization processes, discussed in the third paper, we might have a case where language allows us to make a distinction between different kinds of processes—Type 1 processes are *mostly* nonlinguistic and Type 2 processes are *mostly* linguistic—but, again, this distinction has to be refined. There are very likely more than just two options. Definitions, understood as a cognitive process participating in categorization tasks, are certainly tied to language, but the way they change categorization processes can sometimes be indirect—through redirection of attention towards some features of an animal or an object, for example. Language also cuts across types of processes in this case. Still, this proposal was meant to apply to categorization processes *only*, and not to all cognitive processes. It will be interesting to see, in future work, how such distinctions will work out when studying other cognitive capacities.

But what about the case of Brother John? In 1.1, I presented this case of a French Canadian monk who had long spells of aphasia during which he was still able to accomplish very complex tasks we usually associate with linguistic abilities—such as going to a restaurant, picking up the menu and pointing at something on it with the intention of ordering, e.g., an appetizer. Dennett was not impressed by this argument since his claim is that Brother John's cognition is already contaminated by the memetic viruses of language and, even without active, online, use of language, these linguistic resources can be used.

If we look at this from the perspective of categorization processes, however, we might get a clearer picture if what I argued for in Chapter 4 is right. We would expect—this is only speculation so far but I think it is somewhat intuitive once my arguments are accepted—to see Brother John struggle learning new distinctions that require keeping in mind a complex rule, or having trouble understand a complex distinction that requires relying on a definition. There are some categories he might never be able to learn during his aphasic spells. However, if such a definition was already learned and that his other, largely nonlinguistic, categorization processes had already been trained to make a distinction, we would expect him to be still able to make it. This is well explained within the scaffolding approach, but Dennett's approach allows us to go a step further by giving us a framework to focus on some of the transformative virtues of having definitions as a type of body of information in the conceptual apparatus.

Language, even when not available, might transform the very way we approach the learning of new categories—Carey's (2009) notion of placeholder suggests this. Language, in this sense, does not only *enhance* categorization as it can modify the way it operates and can allow processes to latch on to features that would not be available otherwise. The menu case, I believe, can illustrate this. Brother John had numerous previous experiences with restaurant menus and had a general sense of how a menu, even in a foreign country, would be organized. This is the way Donald phrases the issue: "[Brother John] pointed to a line which he thought might be the hors d'oeuvres [...]" (Donald, 1991, p. 84) Brother John was thus still able to process some categories, think about this category as such (hors d'oeuvres versus entrees), and link this thinking with identifiable groups of items on the menu, which requires processing of the size of fonts and other indications that there are divisions within the list of items on the menu.

From discussions in Chapter 3, we are also warranted in thinking that only using the Type 1 / Type 2 distinction, or similar distinctions, might not be sufficient when designing experiments or in developing models, including those from artificial intelligence (such as in Ashby, Alfonso-Reese, Turken, & Waldron (1998)). In the discussion of dual-process accounts in 4.2.2, I attempted to be as careful as possible with regard to this.

Finally, when it comes to discussing the uniqueness of the human species in the animal realm, Penn, Holyoak, & Povinelli (2008) are skeptical that language, at least on its own, can provide a satisfying explanation. They write that "the spectacular scaffolding provided by language or culture" (Penn et al., 2008, p. 109) cannot explain the difference between human and nonhuman minds in the animal realm. In their article, they take, one after the other, three hypotheses about the role of language in cognition (Penn et al., 2008, pp. 121–123): perspectives where language itself changes the mind (Dennett, 1994, 2009), perspectives where only one component or one feature of language does so (e.g., FLN, cf. Hauser, Chomsky, & Fitch (2002); the introduction of an interface for modules to communicate, cf. Carruthers (2013)), and perspectives where language *drove* the evolution of many other cognitive processes (e.g., Gomila, 2012). They believe the latter is the most plausible. What they deny is that human and nonhuman minds are similar in the way it is usually posited within a Darwinian approach to the evolution of human cognition.

On my proposal, however, their argument would fail since the views they review, as I argued, are not mutually exclusive. It might be the case that we need more than one of these approaches to explain what is going on in human minds that makes them so different from the minds of other species. This point is not decisive against the perspective offered by Penn and colleagues, but I believe it to be a worthwhile avenue to pursue. Bickerton (2014, pp. 263–265) has recently suggested a similar avenue—comparing his view to my own will be the focus of future projects.

It is important to note that the ambition of this project was vast; I never expected to fully develop a theory of the language-cognition interaction in human minds. I am unconvinced that, at this point in time, we understand the language faculty, the architecture of mind and their respective—sometimes shared!—evolution sufficiently. The title of this dissertation was changed accordingly following the defense of the project. In these pages, I offered *remarks* on *many*—I might have added—theories of the language-cognition relationships. My aim was, mainly, to establish that this is a topic of interest that should be the focus of much more philosophical work, viz. research on this topic has not been pursued sufficiently and it should be undertaken. The second goal was to offer examples of different ways in which this research program might be continued. I think this exploratory work has the potential to give us access to new vistas on this complex landscape that—so far—I have just started to explore.

# References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481.

Bickerton, D. (2014). *More than nature needs: language, mind, and evolution.* Cambridge (MA): Harvard University Press.

Botterill, G., & Carruthers, P. (1999). *The philosophy of psychology.* New York: Cambridge University Press.

Carey, S. (2009). *The origin of concepts.* New York: Oxford University Press.

Carruthers, P. (2012). Language in Cognition. In E. Margolis, R. Samuels, & S. P. Stich (Eds.), *The Oxford Handbook of Philosophy of Cognitive Science* (pp. 382–401). New York: Oxford University Press.

Carruthers, P. (2013). The fragmentation of reasoning. In P. Quintanilla (Ed.), *La coevolución de mente y lenguaje: Ontogénesis y filogénesis*. Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú. Retrieved from http://faculty.philosophy.umd.edu/pcarruthers/The%20Fragmentation%20of%20Reasoning.pdf

Churchland, P. M. (2012). *Plato's camera : how the physical brain captures a landscape of abstract universals*. Cambridge (MA): MIT Press.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

Dennett, D. C. (1994). Language and intelligence. In J. Khalfa (Ed.), *What is Intelligence?* (pp. 161–178). Cambridge: Cambridge University Press. Retrieved from http://ase.tufts.edu/cogstud/papers/rolelang.htm

Einstein, A., & Infeld, L. (1938). *The evolution of physics: The growth of ideas from early concepts to relativity and quanta*. New York: Simon & Schuster.

Gawronski, B., & Creighton, L. A. (2013). Dual-Process Theories. In *The Oxford handbook of social cognition*. New York: Oxford University Press. Retrieved from http://publish.uwo.ca/~bgawrons/documents/GC_Oxford.pdf

Gomila, A. (2012). *Verbal minds: language and the architecture of cognition*. Boston: Elsevier.

Machery, E. (2009). *Doing without Concepts*. New York: Oxford University Press.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, *31*(2), 109–130; discussion 130–178. doi:10.1017/S0140525X08003543

# Curriculum Vitae

| | |
|---|---|
| Name | Guillaume Beaulac |
| Post-secondary Education and Degrees | Cégep du Vieux Montréal<br>Montréal, Québec, Canada<br>2003-2005 D.E.C. Histoire et civilisation |
| | Université du Québec à Montréal<br>Montréal, Québec, Canada<br>2006-2008 B.A. Philosophie with Mention d'honneur (2009) |
| | Université du Québec à Montréal<br>Montréal, Québec, Canada<br>2008-2010 M.A. Philosophie with Mention d'honneur (2011) |
| | The University of Western Ontario<br>London, Ontario, Canada<br>2010-2014 Ph.D. Philosophy |
| Major Honors and Awards (over $5,000) | Social Sciences and Humanities Research Council of Canada<br>Joseph-Armand-Bombardier Master's Scholarship<br>2008-2009 |
| | Fonds de Recherche du Québec - Société et culture<br>Bourse de maîtrise en recherche<br>2008-2010 (partly declined) |
| | Fonds de Recherche du Québec - Société et culture<br>Bourse de doctorat en recherche<br>2010-2013 (declined) |
| | Social Sciences and Humanities Research Council of Canada<br>Joseph-Armand-Bombardier Doctoral Scholarship<br>2010-2013 |
| | Province of Ontario Graduate Scholarship<br>2013-2014 |

Richard J. Schmeelk Canada Foundation
Schmeelk Fellowship
2013-2014 (partly declined)

Fonds de Recherche du Québec - Société et culture
Bourse postdoctorale (ranked 1st in Philosophie / Éthique)
2014-2016 (declined)

Social Sciences and Humanities Research Council of Canada
Postdoctoral Fellowship
2014-2016

Related Work Experience

Teaching Assistant
Université du Québec à Montréal
2006-2010

Research Assistant
Université du Québec à Montréal
2007-2008

Assistant to the French Editor of *Dialogue*
Canadian Philosophical Association
2007-2012

Assistant Web Editor
Canadian Philosophical Association
2009-...

Teaching Assistant
University of Western Ontario
2010-2012

Enseignant en philosophie
Cégep Saint-Jean-sur-Richelieu
2013

Chargé de cours (PHI0200 – Pensée critique)
Université du Québec à Montréal
2014

## Publications

(Under review) Beaulac, G., & Poirier, P. “"Concept" Heterogeneity and Definitions”.

(Under review) Benétreau-Dupin, Y., & Beaulac, G. “Fair Numbers. What Data Can and Cannot Tell Us About the Underrepresentation of Women in Philosophy”.

(Under review) Kenyon, T., & Beaulac, G. “Critical Thinking Education and the Scope of Debiasing”.

Maxwell, B., & Beaulac, G. 2013. “The concept of the moral domain in Moral Foundations Theory and Cognitive Developmental Theory: horses for courses?”. *Journal of Moral Education, 42*(3): 360-382.

Beaulac, G., & Robert, S. 2011. “Théories à processus duaux et théories de l’éducation : Le cas de l’enseignement de la pensée critique et de la logique”. *Les ateliers de l’éthique, 6*(1): 63-77.

Poirier, P., & Beaulac, G. 2011. “Le véritable retour des définitions”. *Dialogue, 50*(1): 153-164.

Beaulac, G. 2010. “Wikipedia et la reconnaissance de la spécialisation”. *Cahiers de l’Institut des Sciences Cognitives, 1*: 59-63.

Beaulac, G., & Poirier, P. 2009. “Compte rendu de Va Savoir! De la connaissance en général (Pascal Engel)”. *Dialogue, 48*(1): 217-221.

## Presentations

16 peer-reviewed presentations, 7 presentations in colloquia, 5 invited talks, and 3 poster presentations between 2009 and 2014.