UNIVERSITÉ DU QUÉBEC À MONTRÉAL

A TWO SPEED MIND?
FOR A HEURISTIC INTERPRETATION OF DUAL-PROCESS THEORIES
(L'ESPRIT À DEUX VITESSES ?
POUR UNE INTERPRÉTATION HEURISTIQUE
DES THÉORIES À PROCESSUS DUAUX)

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
À LA MAÎTRISE EN PHILOSOPHIE

PAR
GUILLAUME BEAULAC

DÉCEMBRE 2010

Savoir, et ne point faire usage de ce qu'on sait, c'est pire qu'ignorer.

Alain, *Propos sur l'éducation*

AVANT-PROPOS

Je souhaitais d'abord me consacrer à des questions liées à la philosophie de l'éducation, mais j'ai vite constaté qu'une telle réflexion ne pouvait se détacher de questions liées à la connaissance, à l'apprentissage et à la pédagogie et, plus largement, à des questions touchant aux fondements des sciences cognitives et à de complexes problèmes de nature politique. La démarche de ce mémoire s'inscrit dans un seul de ces aspects, soit les fondements d'une théorie en sciences cognitives. Les autres aspects demeurent toutefois bien présents, comme motivation et, surtout, comme assises de ce travail.

Je tiens à remercier toutes les personnes que j'ai croisées durant mon épopée uqamienne. Je tiens aussi à remercier, plus particulièrement : Pierre Poirier, mon directeur de recherche, professeur de philosophie à l'UQAM, pour ses importantes contributions à l'ensemble du travail réalisé ici, mais aussi pour son rôle comme guide et critique ; Murray Clarke, mon co-directeur, professeur de philosophie à l'Université Concordia ; Luc Faucher et Serge Robert, membres du jury, professeurs de philosophie à l'UQAM ; Normand Baillargeon, professeur au département d'éducation et de pédagogie de l'UQAM ; Mathieu Marion, professeur de philosophie à l'UQAM ainsi que les personnes formidables avec qui j'ai eu un plaisir dingue à organiser la quatrième *La nuit de la philosophie*, Sindy Brodeur, Eve-Lyne Couturier, Marianne Di Croce, Jean-François Landry, Frédéric Legris, Annie Locat, Philippe Marchand et Simon Tremblay-Pepin. Je remercie enfin le CRSH (2008-2009) et le FQRSC (2009-2010) pour le soutien financier dont j'ai pu bénéficier tout au long de mes études de maîtrise.

Merci aussi à celles et celui qui me rappellent qu'il y a de la vie hors des murs de l'université : Roch Beaulac et Linda Charron, mes parents, ainsi que ma soeur Amélie, merci pour le soutien que seule une famille peut offrir. Je dédie ce mémoire à Evelyn McDuff.
Tu l'as subi du début à la fin, c'est la moindre des choses que je puisse t'offrir en échange !

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

RÉSUMÉ

Ce mémoire est consacré aux théories à processus duaux, abondamment discutées dans la littérature récente en sciences cognitives. L'auteur y propose une version fortement amendée de l'approche défendue par Samuels (2009), remplaçant la distinction entre 'Systèmes' par une distinction entre 'Types de processus', qui permet de critiquer à la fois les approches (uniquement) modularistes et les approches décrivant une différence profonde entre deux systèmes ayant chacun leurs spécificités (fonctionnelles, phénoménologiques, neurologiques). Cependant, dans la version des théories à processus duaux défendue ici, la distinction entre 'Types de processus' n'est considérée *a priori* que comme une distinction heuristique permettant aux chercheuses et chercheurs de mieux comprendre l'esprit et d'en expliquer certaines propriétés. L'idée centrale défendue dans ce mémoire est que les processus cognitifs devraient y être distingués selon leur position dans un espace conceptuel multidimensionnel permettant de considérer l'ensemble des caractéristiques et des spécificités attribuées à un processus, cela étant préférable à les forcer dans l'un ou l'autre des 'Systèmes' ou des 'Types' identifiés dans les approches les plus influentes (cf. Evans, 2008). Une fois ce programme de recherche entamé, il sera alors possible de réviser la définition des concepts et des catégories utilisés pour refonder certaines notions présentes dans la littérature ('module', 'Système 1 / 2', etc.). L'argument se déroule en trois temps :

1) Le premier chapitre vise à clarifier la notion de 'module' très utilisée en sciences cognitives. Contre les approches visant à affaiblir cette notion afin que tous les processus de l'esprit soient considérés comme étant des modules, l'auteur – suivant notamment Faucher et Poirier (2009) et Samuels (2006) – jette le doute sur l'emploi qui est fait de cette notion par plusieurs auteurs très influents, notamment en psychologie évolutionniste (par exemple, Barrett, Carruthers, Cosmides, Tooby).

2) L'objectif du second chapitre est de présenter, examiner et critiquer plusieurs théories à processus duaux et de suggérer qu'aucune théorie actuellement discutée n'est adéquate pour décrire l'architecture de l'esprit. Les approches, particulièrement influentes ou représentatives, défendues par Stanovich (1999; 2004; 2009), par Evans (2008; 2009), par Lieberman (2007; 2009) et par Carruthers (2006; 2009) y sont abordées.

3) Dans le troisième chapitre, l'auteur critique le cadre développé par Samuels (2009), puis développe son approche des théories à processus duaux en montrant certains de ses avantages.

Mots clés : philosophie des sciences, sciences cognitives, psychologie, neurosciences, théorie à processus duaux, modularité, module, heuristique, biais, pensée critique, naturalisme.

ABSTRACT

This dissertation is devoted to dual-process theories, widely discussed in the recent literature in cognitive science. The author argues for a significantly modified version of the account suggested by Samuels (2009), replacing the distinction between 'Systems' with a distinction between 'Types of processes,' which allows a critique of both the (only) modularist accounts and the accounts describing a deep difference between two systems each having their specificities (functional, phenomenological and neurological). In the account of dual-process theories developed here, the distinction between 'Types of processes' is considered, *a priori*, only as a heuristic distinction allowing to better understand the mind and to explain some of its properties. The main idea defended in this dissertation is that cognitive processes should be distinguished by their position in a multidimensional conceptual space allowing researchers to consider all of the characteristics and peculiarities attributed to a process, which is preferable to accounts forcing the process into either 'System' or either 'Type' identified in the most influential approaches (cf. Evans, 2008). Once this research program is in place, it will become possible to revise the concepts used and the categories defined to ground anew some of the notions we find in the literature ('module,' 'System 1 / 2,' etc.). The argument is made in three steps:

1) The first chapter aims at a clarification of the notion of 'module' commonly used in cognitive science. Against the accounts aiming at a weakening of the notion in order to classify all of the mind's processes as modular, the author – following most notably Faucher and Poirier (2009) and Samuels (2006) – casts doubt on how this notion is used by many influential authors, particularly in evolutionary psychology (e.g. Barrett, Carruthers, Cosmides, Tooby).

2) The objective of the second chapter is to present, examine and criticize many dual-process theories and to suggest that no theory currently discussed is adequate to describe the architecture of mind. The accounts, especially influent or representative, suggested by Stanovich (1999; 2004; 2009), by Evans (2008; 2009), by Lieberman (2007; 2009) and by Carruthers (2006; 2009), are examined.

3) In the third chapter, the author criticizes the framework developed by Samuels (2009) before developing his own account of dual-process theories and discussing some of its advantages.

Keywords: philosophy of science, cognitive science, psychology, neuroscience, dual-process theory, modularity, module, heuristic, bias, critical thinking, naturalism.

INTRODUCTION

Much of the philosophical tradition conceives the human mind as a unified and rational entity that confers human beings superiority over other animals. This view has been, and is still, tremendously influential in many philosophical debates, including in many accounts of education and of critical thinking.

For instance, in his 2005 book *Petit cours d'autodéfense intellectuelle*, Normand Baillargeon has two goals: to make people aware of the importance of thinking critically, and to teach this skill by offering a guided tour of the diverse skills necessary to achieve the difficult task of thinking critically. Teaching critical thinking, the hardest goal of his book, seems simple and is presented as such: first, we teach logic, notions about denotation and connotation in language, arguments, fallacies, mathematics, statistics, and so on. Then, the students practice these newly acquired skills. The idea of this program is that it will be sufficient for them to develop an acute *esprit critique* and that they will become human beings acting and thinking within the standards of rationality if they wish to do so. In other words, these tools will allow them to be critical thinkers who can avoid being tricked by publicity or fallacies.

The very attractive picture underlying Baillargeon's implementation of his second goal is, sadly, flawed. Moreover, although Baillargeon's implementation, in the end, does not produce the results we would want to see, it is representative of the way critical thinking is usually understood. For example, Bailin and Siegel (2003) suggest there are two aspects to

critical thinking: the ability and the disposition to think critically, the former referring to the acquisition of a skill and the latter, to the virtues possessed by a subject[1].

What these thinkers generally neglect is that acquiring this skill is far more complex than what traditional critical thinking and logic textbooks make us believe. Indeed, rational thinking is hard for human beings; for example, as the heuristics and biases literature shows, it is difficult to apply formal rules or to think along the lines of what is considered rational – even when we know how (and are disposed) to do it (cf. Gilovich, Griffin and Kahneman, 2002). Nevertheless, even though his methods for teaching it look unpromising to contemporary cognitive scientists, Baillargeon's second goal – teaching critical thinking – is a very important one and should be investigated further.

Baillargeon's error, I believe, is that he takes for granted that, with practice, it is easy to apply the rules once we know them. However, many researchers have shown that the problem is not only the lack of knowledge of the rules, but also the lack of knowledge that intuitions lead us astray in some cases: we must realize we need cognitive tools in certain contexts. The idea underlying this dissertation is that not only do we need to properly teach how to apply formal rules to a problem, but we also need to make students aware of how the mechanisms of the mind processes information in order to make it easier for learners to identify and characterize situations in which they must use cognitive tools to override their intuitions (and this insight has some experimental evidence; e.g. Houdé *et al.*, 2000). Stanovich and West (2008) suggested an insightful model of reasoning where they include not only the possession of the correct cognitive tools, but also the importance of knowing when to apply these rules – what they call override detection. Explicitly learning this 'override detection' is only the first step: the students then need to practice and automatize this ability and their use of the appropriate cognitive tools (in the right situation).

---

[1] "The primary disposition consists in valuing good reasoning and being disposed to seek reasons, to assess them, and to govern beliefs and actions on the basis of such assessment. In addition, most theorists outline a subset of dispositions that are also necessary for critical thinking, including open-mindedness, fair-mindedness, independent-mindedness, an inquiring attitude, and respect for others in group inquiry and deliberation." (Bailin and Siegel, 2003, 183)

In order to tackle this hypothesis, and explain why only knowing the rules and how to apply them is not enough, there are some important issues to consider first as we can both understand why we are unable to find the solution to this kind of task easily and how we can learn to use the tools of critical thinking adequately. The goal of this dissertation is to elucidate one important aspect that serves as a foundation for the framework used by, among others, Stanovich and West (2008) to think about these problems and this type of solution. Their hypothesis did not appear *ex nihilo*: their model of reasoning is based upon a family of theories becoming very popular in cognitive science: dual-process theories.

Offering insights into teaching critical thinking is the focal point of my personal overall reflections in cognitive science and the philosophies of mind and education, but what I will discuss here will be in the philosophies of cognitive science and of mind by developing a framework in which we can, I will argue, 'rethink' dual-process theories. A better grasp of cognitive processes should allow us to devise better methods to teach critical thinking. In other words, to see how these mechanisms function we need, first, to understand how the mind works, and this is precisely the goal of this dissertation, which I will achieve by evaluating different frameworks allegedly offering a model of the architecture of mind.

As mentioned, Stanovich and West's (2008) model is based on a specific account of dual-process theories (cf. Stanovich, 2009). These theories are an important family of hypotheses that are helpful to understand cognition, and that can help understanding many problems encountered in heuristics and biases psychology (Kahneman and Frederick, 2002; Stanovich, 2009; Stanovich, Toplak and West, 2008). For instance, they allow us to understand why some biases are systematic. And why they can ultimately be corrected. Yet, while important and influential in cognitive science, these theories have a number of theoretical problems I wish to examine in this dissertation.

I will develop an account of the mind that could be described as a massive modularity hypothesis, influenced by Carruthers' (2006; 2009) view[2]. However, most authors, including Carruthers, assume that there is at least one other type of processes at work: they suggest that there are some features of the mind that are not well captured by the idea of a mind *strictly* composed of 'Fodor-modules' (cf. 1.1 for my definition of this notion), including Fodor himself (cf. Fodor, 2000). Dual-process theorists (e.g. Evans, Sloman and Stanovich may be the first proponents of the contemporary form of dual-process theories; for a review, cf. Evans, 2008; for an historial account, cf. Frankish and Evans, 2009), a form of weak central modularity hypothesis (cf. 1.4), usually speak in terms of System 1 (S1) and System 2 (S2), where S1's processes are evolutionarily old, cognitively-closed, autonomous, automatic and function in parallel, and S2's processes are evolutionarily recent, serial, are not cognitively closed and have the capability of inhibiting the automatic response of S1 (cf. Table 2.2 for more details).

For some, like Carruthers, the S1 / S2 distinction is a good way to think about the processes of the mind, although it does not pick up any deep functional difference[3] between the two systems. Consequently, after discussing the massive modularity view and the basics of dual-process theories, I will examine how Type 2 processes are realized. This question is indeed a matter of some controversy as Carruthers (2006, chapter 4; 2009) views S2 as being realized by multiple iterations of S1's modules, while Lieberman (2007; 2009) suggests that there are two systems, clearly divided (which he calls 'X-system' and 'C-system'), both anatomically / neurologically and functionally (what I will identify as 'deep functional difference'). My own view will be in between, as I will suggest that both Carruthers' and Lieberman's views do not offer a satisfactory explanation of some of the data.

---

[2] This approach to massive modularity is less restrictive than many other accounts; for example, Carruthers (2006, chapter 1) argues that Fodor's (1983) definition is too restrictive, while Samuels (2006) argues that Carruthers' account is *not restrictive enough.*

[3] By 'deep functional difference,' I mean a difference in the functional *architecture* of the system and not merely a difference in how it *functions*.

In short, I will argue, following, then modifying importantly, a suggestion made by Samuels (2009), that a dual-process theory using the notion of 'system' cannot be adequate. I will then suggest we should rather conceive dual-process theories as series of continua of the different characteristics usually attributed to each system (e.g. automatic / controlled, nonverbal / language bound, associative / rule based). But, even if I do not agree with the identification of specific systems (and some of their attributed characteristics, a subject I will not discuss here), I think there are at least two broadly defined *types of processes* at work, being functionally different. However, those are not fixed: Type 2 processes can *become* Type 1 (cf. 3.3.2). This is one reason the current dual-process theories, including Samuels' (2009), could not explain every cognitive process: we need a more comprehensive model to integrate all of the diversity and to explain the complexity of the mind.

0.1 Outline of this dissertation

The ideas behind this dissertation are shaped by many discussions in philosophy and cognitive science about the architecture of mind, especially evolutionary psychology and the psychology of reasoning. The dual-process theories mentioned previously are now widely discussed in this literature and, as Evans (2008) observes, most researchers in cognitive science seem to suppose – at least implicitly – one version or another of this type of theories. Indeed, many examples are present in the cognitive science literature of the last 30 years. In Fodor (1983; 2000), Stanovich (1999; 2004; 2009), Evans (1989; 2008), Sloman (1996; 2002), Samuels (2009), Frankish (2004) and even, to a certain extent, Carruthers (2006; 2009), to name a few from philosophy and from psychology, we can find the idea of a mind having two types of processes, or two systems, to explain how the mind, or certain processes of the mind, works. These theories have different names, and they are not all identical, but the general idea remains: accounts of the mind that are entirely modular or entirely domain-general, i.e. accounts explaining the mind as having only one type of processes, are not explanatorily adequate.

The aim of this dissertation is to defend a particular account of dual-process theories, i.e. an analysis of the mind as composed of different *types of processes*[4] rather than composed of two distinct *systems*[5], or of two groups of processes having predetermined characteristics[6], as most accounts in the literature do (Evans, 2008). I will argue that my account solves many of the problems currently encountered by these dual-process accounts of the mind.

In my first chapter, I will discuss the notion of 'module' widely used in cognitive science as the discussions of modularity in general, but mostly the discussions around the nature of central modularity, offer interesting insights about what these two parts of the mind might be. Some of these discussions about modules are very close to many debates on dual-process theories, especially the nature of S1 or Type 1 processes (moreover, Fodor defends a version of dual-process theories). I will explain Fodor's (1983) modularity of mind thesis and see

---

[4] In this dissertation, the term 'process' is meant as a general term that refers to any causal sequence that produces a certain output (the 'product'). Some processes are not functional, viz. one cannot say that it is their function to produce their outputs (e.g. plate tectonics as a process whose product is the current geological structure of the earth). Many processes are, of course, functional: those made up by humans (e.g. word processing) and those 'designed' by evolution (e.g. digestion or respiration). Of course, the processes of concern here will be *cognitive* processes, viz. processes that manipulate, store and transform information in order to help organisms behave adaptively or intelligently in their environment.

In this context, a 'type of processes' refers to a set of processes grouped together because they share one or more characteristic that make them theoretically or explanatorily relevant: while each process is a token, some authors (e.g. Evans, 2008; Samuels, 2009; Stanovich, 2009) argue it is possible to regroup some of these processes into either one of two types – associated with the characteristics usually attributed to System 1 or System 2.

[5] A 'system' is any causally interconnected set of components, from which interaction emerges (as described by Wimsatt, 1986) a property or capacity (the systemic property or capacity). The parts are interconnected the way they are and the systems have these properties they have for causal reasons; for example, in most frameworks under scrutiny here, they have these properties for evolutionary reasons. Systems have a specific structure (their parts and the way they are interconnected) and a certain behavior (way(s) of processing information). Of concern here the dual emergent properties of types of cognitive (sub)systems, dubbed 'System 1' (automatic, fast, parallel, etc.) and 'System 2' (controlled, slow, serial, etc.). In 2.1, I will describe how dual-process theories use the notion of 'system' in more details.

[6] This could be the case if the S1 / S2 or Type 1 / Type 2 distinction identified natural kinds. I hope to show we currently do not have sufficient evidence to conclude that they are.

how this is an account of dual-process theory where central cognition is not (at all) modular. Then I will explore how this notion of 'module' has been used by evolutionary psychologists to explain parts of central cognition. Next, I will present a recent, more encompassing[7], account of central modularity, where the mind is entirely modular, defended most notably by Carruthers (2006, chapter 1). I will finally introduce some of Carruthers' critics and defend Samuels' (2006) weak central modularity hypothesis.

In my second chapter, I will discuss the basics of dual-process theories and argue that one consequence of the debates around massive modularity is to adopt one type or another of dual-process account of the mind, even for its most ardent defenders[8]. However, Carruthers (2009) takes another path by suggesting that S2 results from the multiple iterations of capacities implemented in S1 systems. For him, there is no deep functional difference between the S1 and S2, there are only differences in how the output of a system *feels like* to the organism that has the system and how the system is realized. For Carruthers, explaining S2 as multiple iterations of capacities implemented in S1 systems implies that there are no deep functional differences between a cognitive system with only S1 capacities and one with S2 capacities, but there *are* differences in how it works (these multiple iterations are different than a single iteration of one of S1's capacities). Moreover, it is this '(shallow) functional difference,' how a given process works, what gives rise to phenomenological differences. Carruthers' version of the S1 versus S2 distinction will be contrasted with Lieberman's position: Lieberman believes there are deep functional differences between S1 and S2 (Cohen and Lieberman, 2010; Lieberman, 2007; 2009; Satpute and Lieberman, 2006; see also its presentation by Evans, 2008, 270). He believes that S1 and S2 are well-defined systems we can clearly identify to specific neuronal activation patterns, and goes as far as to attribute some cognitive capacities to specific brain

---

[7] The characteristics Carruthers attributes to modules are weaker, which makes his account – a strong central modularity thesis – a more encompassing account of the modularity of mind.

[8] I will also highlight that the weak central modularity thesis is more plausible than the 'strong' version.

regions. Both Carruthers' and Lieberman's account are unsatisfying as they fall short of explaining a number of cognitive processes. Still, other dual-process theorists are not more successful in tempting to describe and explain processes that do not fall in either S1 or S2 – I will be interested in accounts defended by Stanovich (2009) and by Evans (2009).

In my third chapter, I will develop my own account. *Contra* Carruthers and Lieberman, I will suggest we should cease to use the notion of 'System' altogether when talking about this difference (more or less in line with Samuels', 2009 suggestion), and instead speak of Type 1 and Type 2 *processes*. However, the characteristics attributed to each 'Type' do not align as Samuels (2009) pretends, viz. the dichotomies identified in Table 2.2 do not necessarily co-vary and it is possible to find, for example, automatic and controlled processes or others that are nonverbal and conscious: for this reason we need a framework where we can take this important fact into account. In a nutshell, my view is that to understand the mind, we need a series of continua, one for each of the different characteristics posited by dual-process theorists, thereby creating a n-dimensional conceptual space (where n is the number of characteristics considered) we can use to situate every cognitive process according to its own properties. The Type 1 / Type 2 division would thus simply be a shortcut to locate regions in this conceptual space where most processes cluster when their characteristic are plotted in the space.

After arguing for this view, I will discuss its three main advantages: it allows us (i) to posit complex processes having both Type 1 and Type 2 characteristics (such as one discussed by Evans, 2009, but I will provide other examples), (ii) to account, conceptually at least, for the possibility of processes becoming *modularized*, i.e. for a Type 2 process to acquire some (and even most) characteristics of Type 1 processes (exploring thereby an idea suggested, among others, by Karmiloff-Smith, 1992 and also providing a framework for Dreyfus and Dreyfus', 1986 observations) and, finally, the conceptual space my view introduces (iii) could be very useful for further research.

In the end, this emphasis makes this dissertation an essay in the philosophy of cognitive science, especially in the philosophy of psychology, more than one concerning the philosophy of education. I strongly believe that the discussion presented more generally in

this dissertation should not be neglected if we are to make useful recommendations about what, when, how and why we are to teach children (or adults!) as knowing more about the architecture of the mind can only help us in devising new and better tools.

0.2 Viewing the mind evolutionarily: the framework

The argument developed in this dissertation relies, first and foremost, on an evolutionary framework. The human body is seen as the result of a long and complex evolutionary process, and this claim is not very controversial. For example, some features of the human body are clear examples of *bad* (i.e. not optimal) design (retinal blind spot, spine not ideal for upright position, esophagus can cause choking, and so on). On this basis, the idea of the mind as a unified and rational entity has been challenged, for instance by Marcus (2008). For him, not only the human body is a *kluge*: the human mind also is one. A kluge, in the vocabulary of engineers, is a functioning but "clumsy or inelegant" (Marcus, 2008, 2) solution to a complex problem.

In the case of the mind, its *engineer*, the process of natural selection over small genetic mutations, has no goals, *a fortiori* no goal such as an optimal or elegant solution: only survival and reproduction matter to her. If an inelegant solution is efficient, it will be passed on. Moreover, as Jacob (1970) observed, natural selection can only work with what it has at hand: it cannot redesign a system from scratch. Therefore, the mind is the product of an evolutionary development *that always occurs on top of what was already there*: Marcus talks about layers of 'technology,' with the upper layer working with what is in place in the layers underneath. The mind is then viewed as the product of an evolutionary process and, this is a key to evolutionary explanations as Dawkins (1986) argues. The structures or capacities added might allow for more efficient process, but these processes might be limited (or, at least, limited in certain cases or in certain environments). When studying the mind, understood as the result of an evolutionary process, we must then question the assumption that a given cognitive process is optimal, or even that the process in question is perfectly adapted to its task (as it could be only a by-product, or it could only be partially adapted): it

must be proven. Additionally, an elegant architecture cannot be presupposed: such a discovery is possible (but improbable) but, here again, it cannot be taken for granted.

Such a view is important because it offers a new insight on what seems to be limitations of our mind discussed by psychologists, such as Wason, Evans, Kahneman, Tversky and many others (e.g. Bishop and Trout, 2005; Cosmides and Tooby, 2000; Gilovich *et al.*, 2002; Marcus, 2008; Stanovich, 1999; 2004; 2009; Stanovich *et al.*, 2008; Stanovich and West, 2008). Indeed, this view of the mind as a kluge is in line with the evolutionary approaches in psychology by holding that cognition is the product of evolution. The mind is made, at least partly, of many modules; as Gigerenzer, Todd and the ABC Research Group (1999) suggested. It is an assembly of systems functioning with simple heuristics that are good enough in ecological environments, but that produce suboptimal outputs in some situations[9].

In the literature, such a framework has appeared under the label of 'evolutionary psychology,' and these researchers made much progress to understand the characteristics of the human mind. In practice however, their program focuses on the idea of modules in a more or less Fodorian sense, often without elaborating on other types of processes (cf. 1.2). Many researchers question this view and explain cognition with a more flexible framework, and it is at the center of this dissertation. Dual-process theories are indeed an alternative much discussed in the literature (Carruthers, 2006; Evans and Frankish, 2009; Stanovich, 2004; 2009) and my goal is to offer an account that could potentially solve most of its current difficulties (as identified by e.g. Evans, 2008; Keren and Schul, 2009; Machery, forthcoming, section 8; Samuels, 2009) without losing its explanatory power.

---

[9] Gigerenzer (2007) seems to argue that these situations are scarce; I do not agree with this position since the environment in which these heuristics are activated is now very different from the environment in which this heuristic evolved. It does not mean in any way that these heuristics lost their usefulness or their accuracy in *a great deal* of contexts; it only means that, in many situations, we cannot rely on these natural heuristics alone. There are many cultural heuristics (meta-heuristics, cf. Wimsatt, 2007) that can help us when these automatic processes go astray (cf. Afterword).

CHAPTER I

DOES THE MIND WORK THAT WAY? MODULARITY AND ITS LIMITS

What exactly is a module and what does the modularity hypothesis explain is subject to a large amount of controversy. I will begin by presenting Fodor's (1983) account of modules before assessing what modularity can and cannot explain (Fodor, 2000) by looking at discussions about modules in the evolutionary psychology literature. Fodor's influential proposal was interpreted in diverse ways, but the most discussed of these is most probably the interpretation of modularity made by Evolutionary Psychologists[10].

Their suggestion is that central cognition is, at least partly, modular. They introduced, with the massive modularity hypotheses, what Samuels (2006) called central modularity hypotheses, and which exist in different degrees:

> *Strong central modularity*: All central systems are domain specific and/or encapsulated, and there are a great many of them.
>
> *Weak central modularity*: There are a number of domain specific and/or encapsulated central systems, but there are also nonmodular – domain general and unencapsulated – central systems as well. (Samuels, 2006, 42)

---

[10] Here, I am mostly referring to the so-called 'Santa Barbara evolutionary psychology school' around Cosmides and Tooby, those who Buller calls Evolutionary Psychologists (capital 'E' and 'P,' cf. Buller, 2005).

This distinction is essential to understand the debates over the nature of modularity. Fodor's (1983) account of modularity is different from that of Evolutionary Psychologists on many grounds (cf. Faucher and Poirier, 2009), and this is a crucial point. Many Evolutionary Psychologists argue that the mind is massively modular, i.e. that most of cognition, including a large part of central cognition, is modular. It is important to note that, although there might be literally thousands of modules according to some of them, they are not suggesting that there is *nothing else besides* modules (cf. Cosmides and Tooby, 1994; Clarke, 2004, chapter 1). At least a part of central cognition is modular: they defend an account of weak central modularity.

In the literature on modules however, there are some who defend strong central modularity hypotheses, such as Barrett (2009; Barrett and Kurzban, 2006) and Carruthers (2006; 2009). For them, as Samuels explains, there is nothing other than modules that make up the mind. While this might seem implausible given Fodor's characterization of modules (cf. 1.1), their strategy – and the initial strategy adopted by Evolutionary Psychologists – is to redefine what 'module' means by removing some of the characteristics Fodor ascribed to them. The general idea is that weaker notions of module can explain some aspects of central cognition, up to the point where, even in central cognition, there is nothing else besides modules.

For his part, Fodor insists that modularity can only explain peripheral cognition, viz. that it cannot explain any aspect of central cognition (e.g. our capacity to do abduction). In other words, for Fodor (1983; 2000), central cognition is not modular to any extent.

After explaining what Fodor means by 'module,' I will explore the strong central modularity hypotheses of Barrett (and some Evolutionary Psychologists) and Carruthers. I will then criticize Carruthers' account, and suggest that dual-process theories might answer many of the problems of massively modular accounts of the mind.

1.1 Fodor's modularity

Fodor-modules have nine properties: their operation is domain specific, mandatory once activated (and they are automatic), fast, informationally encapsulated, their processes are not centrally accessible (only their output is), they have shallow outputs, exhibit specific 'breakdown patterns' and characteristic pace and sequencing (they have an ontogenetic timetable, viz. specific developmental characteristics), and finally they are associated with a fixed neural architecture.

In this dissertation, I will use the notion of 'Fodor-module' to refer to any module possessing all these nine characteristics. However, it is crucial to remember that Fodor states explicitly that modules may miss some of these characteristics, as it will be critical when I will discuss dual-process theories in Chapter II. In other words, Fodor is not explicitly defending a dual-process account where there are only Fodor-modules and central / higher cognition; he might allow some modules with only some of the nine characteristics presented in the preceding paragraph. Furthermore, my goal is not to present each of these characteristics in detail (especially since some of these are linked – possibly causally – to others) but only some of them, those that are important to the thesis I will develop in this dissertation.

The idea I want to emphasize below is simple: Fodor's account is interesting to explain *some* cognitive processes, but certainly not all of them. Of course, he limits his account to peripheral cognition[11], mostly input systems, by specifying that "[e]ven if input systems are domain specific, there must be some cognitive mechanisms that are not" (Fodor, 1983, 101), but even those input systems, as I will argue, might be more diverse than his strong modularity account allows, viz. some input systems exhibit some *non*-modular traits (and many central processes or systems have modular traits).

---

[11] As Prinz (2006, 22) notes: "The book would have been more aptly, if less provocatively, called *The Modularity of Low-Level Peripheral Systems.*"

One of the best known and most discussed characteristic of modules is their domain specificity. Modules, according to Fodor, are specialized to respond to certain inputs, and usually have their own sensory transducers. They are limited to a particular type of inputs, and this would be why they are as efficient as they are[12]. Understood in this way, domain specificity is problematic on a number of levels.

The way to determine how many modules there are in the mind depends mostly on what is considered to be a domain, i.e. it depends on the grain used to analyze, and it also depends on what modules are considered to be (Can we identify and distinguish modules functionally?, Are modules natural kinds?, etc.). And, once the proper grain is found (if there is a way to do so; cf. Prinz, 2006, 27-30 for a sceptical outlook), the jury might still be out to find the proper description of a given process (Atkinson and Wheeler, 2004). Vision, for example, likely needs an important number of modules to do its task, but should we count horizontal line detectors and vertical line detectors as two types of modules, or is it more interesting to use the more general class of *edge detectors*? The description of the visual system will likely change in important ways depending of the answer we give here. The general idea still remains important: a module answers to (or is activated by) a certain classes of inputs, and determining how to find the proper way and the correct level of analysis is largely up to empirical research[13].

Modules are automatic and mandatory: visual or auditory illusions provide a good example of this characteristic. If the appropriate stimulus is presented and seen or heard, the illusion will be seen or heard (the McGurk effect, cf. McGurk and Macdonald, 1976, is a nice

---

[12] An Evolutionary Psychologist would say the input system *evolved for* solving an evolutionary problem, and this is why a given module is specialized (cf. Carruthers, 2006, chapter 1) an idea Fodor would probably try to resist (cf. Fodor, 2000; Fodor and Piattelli-Palmarini, 2010).

[13] For instance, Machery (2009) argues that 'concept' is not a natural kind, i.e. it is not a useful generalization in psychology. For him, there are different kinds of structure, different types of modules we might say, that are used to do the different tasks traditionally attributed to a more general process. Machery's contribution is to show that there are very good reasons to distinguish between 'prototype-concepts,' 'exemplar-concepts' and 'theory-concepts' and, for him, they are different modular processes.

example of senses combination). Fodor presents three examples to illustrate his idea: first, if someone hears a given utterance in a known language, she will hear a sentence and give it meaning; second, any object perceived is perceived in a three-dimensional space; third, touching a surface entails feeling it. The first example he gives, as we will see, might be problematic (cf. 3.3.2), but this characteristic is otherwise pretty straight forward: it is not possible for an unimpaired subject not to see when his eyes are opened, not to feel when he touches an object, not to hear when a noise or not to taste or smell when certain molecules come in contact with the appropriate receptors, and the same goes for other types of higher level modules such as speech recognition (if I hear French, I will attribute meaning to the utterance).

Modules are fast: once activated, a module usually produces its output well under a quarter of a second. The two important aspects this reveals according to Fodor (1983, 63) are the contrast between the speed of the modules' processes as opposed to how slow central processes can be, and the strong link between speed and their mandatory operation.

As Fodor explains, some of the computational problems we solve automatically, such as being able to identify an object against its background, are not necessarily easier or harder computational problems *per se* than solving a long mathematical equation[14]. He even goes further by suggesting that "[t]his dissimilarity between perception and thought is surely so adequately robust that it is unlikely to be an artefact of the way that we individuate cognitive achievements" (Fodor, 1983, 63), something Carruthers disagrees with (cf. 2.2). The second idea, that "processes of input analysis are fast *because* they are mandatory" (Fodor, 1983, 64), is related to the fact that an automatic process does not involve any

[14] The first case can be extremely difficult to implement in artificial agents, but it is easy to do for human beings; the second case is harder for human beings than for artificial agents.

decision making process[15]. Reflexes are faster than deciding to do a given action because, simply put, "making your mind up takes time" (Fodor, 1983, 64).

The outputs of Fodor-modules are shallow. The notion of shallowness, as applied to modules, is ambiguous, both in Fodor (1983, 86-97) and in the literature more generally. He sees it as the difference between "observation and inference" (1983, 86), shallow outputs being those directly observable whereas non-shallow (deep?) outputs being those that can be inferred (e.g. we can directly observe that a traffic signal is red but can only infer that it is meant to signal the obligation of stopping). However, he remains vague as to exactly what he means by this, and Fodor's vagueness has given rise to a number of distinct interpretations.

Prinz, for instance, rejects the idea that "[s]hallow outputs are outputs that do not require a lot of processing" (2006, 25) because it is not precise enough to create any meaningful categorization (what is 'a lot' is not clear). Carruthers prefers to understand the idea that outputs of modules are shallow as meaning that they are nonconceptual (Carruthers, 2006, 4), and this is certainly a strange interpretation of shallowness (e.g. Fodor would probably not agree with this idea, as even the most basic categories, the primitives, are 'concepts' for him; cf. Fodor, 1998, chapter 2). Faucher and Poirier prefer to explain shallowness in terms of 'basic categories,' viz. categories that do not use background knowledge, categories that are simpler representations (Faucher and Poirier, 2009, 287). This view is less controversial, and more in line with Fodor's project generally (e.g. Fodor, 1998). This means that we should understand shallowness as meaning that the modules' outputs are not theoretically charged which seems an interesting way to understand modules' shallowness in the contemporary context. By combining aspects of the three views discussed above, we obtain an account of shallowness as the fact that the outputs of modules involve no reconceptualization using background knowledge because the processes are strongly encapsulated, and this absence of reconceptualization is the reason the outputs of module do not require much processing.

---

[15] "Because these processes are automatic, you save computation (hence time) that would otherwise have to be devoted to deciding whether, and how, they ought to be performed." (Fodor, 1983, 64)

Modular processes are not centrally accessible, viz. we do not have introspective knowledge of their workings. Only their outputs are accessible to other modules or to central cognition. This latter characteristic is analogous to informational encapsulation: the inaccessibility characteristic specifies that the representations within a module are not accessible to central processes, while informational encapsulation tells us that modules cannot have (direct) access to the content of central cognition (or that of other modules). For example, visual modules do not have access to our knowledge that a given picture is an optical illusion (and we cannot consciously affect the inner workings of the module), hence the illusion persists and there is no *direct* way of affecting it.

Fodor argues modules might be associated with fixed neural architectures (Fodor, 1983, 98), and that these modules also exhibit specific ontogenetic sequencing. Indeed: during development, some capacities appear at specific moments and in characteristic ways. Language acquisition is the best known case, but there seems to be such developmental constancy in folk physics, folk psychology, mathematics, etc. Piaget's work is the best known on some of these questions and, even if his framework is opposed to domain specificity, the neopiagetians' framework is not (e.g. Gopnik and Meltzoff, 1996; Gopnik, Meltzoff and Kuhl, 2000[16]).

As I already mentioned, many processes have these characteristics according to Fodor, but the modularity of mind applies only to peripheral cognition. He argues, and insists on this idea, that central processes are not and could not be modular. For example, central processes are argued to be quinean, viz. processes of the central system are taken to be potentially sensitive to the whole set of beliefs held by the subject (holism), making these processes unencapsulated. Fodor's account is widely debated and many disagree with his view, stating it is too restrictive (Carruthers, 2006; Samuels, 2006), sometimes going as far as rejecting modularity altogether (Prinz, 2006).

---

[16] Another example is Karmiloff-Smith's (1992) work, where she offers an interesting perspective on modularity, different from Piaget's and Fodor's, which I will discuss in 3.3.2.

Another account of modularity is however possible: Evolutionary Psychologists began to argue for the massive modularity hypothesis, in which even central cognition might be, at least partially, modular, an idea against which Fodor argued in his 2000 book. An important point is however rarely mentioned: many of these debates are about how to define 'module' rather than about modularity itself.

1.2 Evolutionary Psychology, massive modularity and the many definitions of 'module'

Evolutionary Psychologists use the notion of module in the framework of the massive modularity hypothesis. The idea, against Fodor's (1983, 2000), is that there is more to 'the modularity of mind' than just an explanation of the peripheral systems. Most Evolutionary Psychologists adopt one form or another of a central modularity hypothesis. Here we must appeal to the distinction made between strong and weak accounts of central modularity at the beginning of this chapter. In this section and in the next, I will discuss the strong central modularity account before suggesting we adopt a weaker central modularity hypothesis in 1.4.

If, in many cases, the weak central modularity thesis seems to be preferred to the more restrictive accounts (e.g. Clarke, 2004, chapter 1), according to Barrett, the *only* essential characteristic we should consider when defining 'module' is functional specificity (Barrett, 2009, 779). If Barrett's definition is accepted, any process of the mind, if it has one (or more) specific function, will be counted as a module, without further consideration as to how it works. This, of course, might be seen as problematic since the notion of module would thereby lose its very substance: 'module' would then become an uncontroversial notion that does not add anything to traditional 'boxology' in psychology (Faucher and Poirier, 2009).

The problem encountered here is the difficulty of defining what a 'module' is. It remains hard to define this term as it is central to many theories, and this difficulty appears clearly with Barrett's minimal characterization.

This notion is used to talk about a large number of cognitive phenomenon and "not only have authors used the term *modular* to refer to different concepts, but even explicit definition of the term by some researchers has been insufficient to avoid subsequent misunderstandings by others" (Barrett and Kurzban, 2006, 642). Moreover, there seems to be "no agreement on a workable characterization of modules for evolutionary psychology" (Downes, 2008, section 3). Still, in all instances, modules are seen by Evolutionary Psychologists as "adaptations, specific to a domain most of the time" (Faucher and Poirier, 2009, 296, my translation). They were selected, through natural selection, for solving adaptive problems such as detecting cheaters in social exchange and recognize kin (these are some of the most common and discussed examples).

If the view of modules as merely adaptations were generally agreed upon by Evolutionary Psychologists, the notion of module, as defined in this functional account, would lack the explanatory power of a richer notion of module. A very general characterization of module will encompass many distinct processes, but, once we identify one such process as a module, we will have very few details on the process in question. A richer account however would provide us with much more information once a module is identified. The analogy I have in mind is the following: in chemistry, if we identify a substance as a non-metal, we will be able to know some of its characteristics, but if we identify the same substance as oxygen, which has a richer definition (including its precise chemical structure), we will know a good deal more about the substance in question.

Of course, as Barrett and Kurzban indicate, modularity in general can help "direc[t] the search for specialization," especially in a framework where evolution has a role to play, as it "constrains the hypothesis space regarding plausible functions" (2006, 643). Yet, this notion is far from the one we began this discussion with and it has nothing left in common with Fodor-modules (or even Fodor's account more generally).

These various changes in the definition of the notion of 'module' used by Evolutionary Psychologists might be mostly explained by diverse debates that occurred around Fodor's *The Mind Doesn't Work That Way*, a clear attack on Pinker's (1997) massive modularity hypothesis in the context of Evolutionary Psychology, but also around critical studies of the

underlying principles of Evolutionary Psychologists' theories and their assumptions (e.g. Buller, 2005; Samuels, 1998). The original research program put forward by Tooby and Cosmides (1992[17]) was ambiguous, to say the least, about what modules are and this lead to a series of discussions culminating in a complete dissociation with Fodor's (1983) notion in order to adopt a notion closer to Barrett's (2009):

> [...] Fodor's (1983) concept of a module is neither useful nor important for evolutionary psychologists. For evolutionary psychologists, the original sense of module – a program organized to perform a particular function is the correct one, but with an evolutionary twist on the concept of function. (Ermer, Cosmides and Tooby, 2007, 153)

In other words, on this account 'module' does not mean much more than a process that has a particular function. If it is the case, and evidence suggests it is, Faucher and Poirier propose to Evolutionary Psychologists that they should "simply stop talking about *massive modularity*, and rather talk in terms of a mind massively constituted of *adaptive structures*" (2009, 307, my translation and emphasis) because the mainstream use of 'module' in cognitive science still refers to Fodor-modules or similar structures (with less and / or modified characteristics). Modules, as defined by Barrett (2009) and by Ermer *et al.* (2007), have a lot less explanatory power and this is an important loss if we are to explain how the mind works.

Carruthers however has developed an interesting account of modules, where he develops another kind of strong central modularity to defend Evolutionary Psychologists against Fodor's arguments but does not do so by removing all substance to the notion of 'module,' as is the case with Barrett.

---

[17] While this article in particular, and more generally, the 1992 book (Barkow, Cosmides and Tooby, 1992), might not be the absolute first presentation of the Evolutionary Psychologists' framework, it is arguably its first detailed account, and certainly the most influential one (e.g. over 2000 citations according to *Google Scholar*). The earliest presentation of their position is most probably Cosmides and Tooby's 1987 article. The primer published online (Cosmides and Tooby, 2000) is a more recent presentation of this same research program, where they (still) remain vague about what 'module' refers to, although they explicitly refer to Fodor-modules.

1.3 Carruthers' account of massive modularity: strong central modularity and weak modules

Fodor is clear: we should not (and we could not) understand the mind as being *only* modular[18]. Nonetheless, Carruthers (2006; 2009) suggests that, with a weaker account of what a module is, we can have a framework where the mind is only composed of modules. Of course, he agrees that peripheral cognition is modular, but he goes further by arguing that central cognition also is entirely modular. A major difference between Carruthers' approach and many of the accounts previously discussed is that he argues for massive modularity, without relying on an insubstantial characterization such as Barrett's, clearly identifying many traits of modules. Moreover, he defends Evolutionary Psychologists by providing a philosophically plausible approach to modularity.

His account is related to the strong central modularity thesis; in fact, in his massive modularity account, *all* cognitive processes are either modular or emerge from the interaction of modular processes. To do so in a plausible way, he weakens the notion of 'Fodor-module' by removing some of its characteristics and redefining others. In fact, even if his notion is richer than Barrett's (2009), the notion he ends up with is so inclusive that Samuels (2006) and Prinz (2006) are not convinced it could be of much use (cf. 1.4).

Briefly, for Carruthers, modules are processing systems, usually associated with a functional domain, that are frugal in their operations and are more or less strongly encapsulated (he introduces the *wide-scope* versus *narrow-scope* distinction, which I explain below), and by and large, only the outputs of a modular process will be available to other processes (Carruthers, 2006, 62-63). However, his account is mostly liberal and allows a lot of variability in each of the characteristics attributed to modules.

---

[18] As a reminder: Fodor (1983; 2000) suggests there *needs* to be non-modular central processes.

Carruthers begins by rejecting some of the characteristics of Fodor-modules, since they would be incompatible with *any* account of central modularity. The shallowness of the output is the first rejected by Carruthers, but he also discards speed. Following these modifications, he takes as plausible domain-specificity, the mandatory and innate character of modules and the neural specificity characteristics, before modifying what is meant by encapsulation and, then, adding frugality.

Carruthers rejects the shallowness of modules' outputs for an obscure reason. He states that if the mind is massively modular as the strong central modularists posit, then there will have to be 'conceptual modules[19].' Then, it will be necessary to explain the outputs of 'conceptual modules,' since only such outputs can provide the "fully conceptual thoughts or beliefs" (Carruthers, 2006, 8) conceptual modules would produce as outputs. He is not clear as to why this is implied: we could easily imagine conceptually shallow content related to a module Carruthers identifies as conceptual, such as kin recognition: recognizing as kin those you have seen for long periods of time during childhood can be done 'shallowly.' I believe the same goes for more complex outputs. Of course, ultimately, conceptual content would have to come into the picture, but Carruthers has, I believe, to have a solution for this in his own framework (cf. 2.2). Of course, if we understand shallowness in terms of nonconceptuality (Carruthers, 2006, 4), Carruthers' point can be made easily because it then becomes somewhat absurd to characterize all modules as shallow, and such shallowness would not even be compatible with Fodor's own account of what a concept is (cf. Fodor, 1998).

The rejection of 'speed' as a characteristic is clearer: modules are fast, according to Fodor, but this characterization only makes sense when the modules' speed is compared to the speed of central processes, viz. modules are faster than central cognition. Without any such comparison, since both peripheral and central cognition are entirely modular for

---

[19] Conceptual modules are, for Carruthers, modules dealing with common-sense physics or biology, or modules used for kin recognition or cheater-detection in social exchanges. As I pointed out previously, mentions of these modules are the most commonly found in the literature.

Carruthers, it makes little sense of maintaining that modules are fast (or slow) (Carruthers, 2006, 9).

At first sight, it seems domain specificity might also need to go because, in a massively modular architecture, some modules would need to receive inputs of all kinds: Carruthers' example is practical reasoning (cf. Carruthers, 2004), and such a module could not be 'domain specific.' In this context, one could understand domain specificity in functional terms (modular processes are restricted to a domain) instead of in terms of content (modules have domain specific content but domain-general processes) as it is usually done[20]. We would obtain a notion of modularity similar to Barrett's (2009), but more precise (as it possesses other characteristics). Although Carruthers agrees modules are systems with "a distinctive function, or set of functions" (2006, 62), he prefers to keep domain specificity and he states that it is possible to understand practical reasoning as being "underpinned by a whole host of different [modular] systems" (Carruthers, 2006, 8), each of which would be domain specific. Still, he is not firm on the necessity of this characteristic: while most modules are likely to be domain specific, some of them could be domain-general (or *more* general, similar as what he suggests for encapsulation, as discussed below).

Carruthers sees some modules, even central ones, as being mandatory since they will automatically process any input they receive. Just as it is the case for the perception of one line as being longer than the other in the Müller-Lyer illusion, we cannot 'turn off' most of our faculties. For Carruthers, "most (if not all) of the component systems that make up the human mind are mandatory in this sense [they can't be *turned off at will*]" (2006, 9), but it also leaves the possibility that, for *some* of the components (probably just a few), it could be possible to interrupt their operation. The example he provides for the mandatory character of 'higher' modules is mind-reading: seeing an actor being sad on stage gives the impression he is sad, even if we know he is not.

---

[20] Content domain specificity allows Samuels (1998) to argue for the plausibility of the library model of cognition (LMC; cf. Carruthers, 2006, section 4.3 for a discussion of the LMC).

Innateness and neural specificity are controversial characteristics compatible with Carruthers' massive modularity hypothesis, but they are not central to his hypothesis. While he tends towards, as he says, "the nativist end of the spectrum" (Carruthers, 2006, 10), this position is not necessary for a coherent massive modularity hypothesis on his view. Neural specificity might be more important, and easier to support empirically as decades of work in neuroscience showed that many functions are implemented in particular areas or have particular pathways in the brain. There are also some important constancies from brain to brain (such as the way eyes are wired to the occipital lobe), although the plasticity of the brain is one of its most important features (Buller, 2005). The idea of the plasticity of the brain can inform massive modularity by specifying that some modules are not 'hard-wired' but developed through specific attention biases in early childhood (Karmiloff-Smith, 1992). Still, one could accept these ideas and agree with Carruthers' massive modularity.

Without firm (content) domain specificity and strong encapsulation, an old problem arises for Carruthers' version of the massive modularity hypothesis: the frame problem, a problem we necessarily need to solve to have a plausible account of the mind. Briefly, the frame problem is the difficulty in AI to limit the information that a system has to consider before initiating an action, and to limit the information it has to update once the action is accomplished. The best way not to be overwhelmed by all the options available to a system (here: a module), both natural and artificial, is, of course, to give its processes a subset of all possible inputs to compute. That is, cognitive processes must be computationally frugal. Encapsulation might be the easiest way to ensure having optimal computationally frugal processes, but Carruthers suggest a weaker version of this important characteristic (Carruthers, 2006, 57). Indeed, as I discussed in introduction (cf. section 0.2), in an evolutionary framework, the goal is survival and not the production of optimal outputs. Reliable and efficient ones will suffice.

Encapsulation, for Carruthers, can be either narrow or wide-scope. Narrow-scope encapsulation means that "concerning most of the information held in the mind, the system in question *can't* be affected by *that* information in the course of processing" (Carruthers, 2006, 58). Wide-scope encapsulation, on the other hand, means that "the system is such that it *can't* be affected by *most* of the information held in the mind in the course of

its processing" (Carruthers, 2006, 58). The difference here is simply one of degree[21]: narrow-scope encapsulation is more encompassing, and it is the notion typically discussed when it is question of modularity. Carruthers argues that wide-scope encapsulation will suffice for his notion of modularity, mostly because he believes there is another way to achieve computational frugality than using narrow-scope encapsulation: heuristics.

In very complex situations where the flow of information is too great, having simple rules of thumb can help take the best possible decision or make the right choice most of the time (Simon, 1957). Of course, these rules of thumb may not always work, they "have been designed to be *good enough*[22]" (Carruthers, 2006, 54; for a discussion see Gigerenzer, Czerlinski and Martignon, 2002); they usually are reliable and efficient. According to Carruthers, inspired by the ideas Gigerenzer uses[23] (e.g. Gigerenzer, 2007; Gigerenzer *et al.*, 1999), this is an alternate solution to the frame problem, different than encapsulation (and sometimes more plausible). The best known example of what a heuristic is the recognition heuristic (cf. Gigerenzer *et al.*, 1999, chapter 2). In a situation when one has to tell which of two cities is the more populous, choosing the city whose name you know, if there is just one, is usually a good way to give the right answer. In fact, when choosing which of two German cities is the most populous, German students fared poorly compared to

---

[21] In logical formulation, for the narrow-scope encapsulation, the negation is after the universal operator, and it is before the universal operator in the wide-scope encapsulation. Where i stands for "information outside of the module's domain" and A(x) for "The system is affected by":
Narrow-scope: $\forall i \neg A(i)$
(For all information outside of the module's domain, the system is not affected by it.)
Wide-scope: $\neg\forall i\ A(i) \equiv \exists i \neg A(i)$
(Not for all information outside of the module's domain, the system is affected by it, and it is equivalent to there exists information outside of the module's domain that does not affect the system.)

[22] Carruthers implicitly refers to *satisficing*, a term coined by Simon (1957). Gigerenzer uses this computer science term in an evolutionary context (evolutionary processes seek satisficing, not maximizing or optimizing), but the idea behind is the same.

[23] The idea was established in computer science by Newell and Simon (1976, section 2). Just as it is the case with Simon's notion of satisficing, Gigerenzer adds an evolutionary flavor to the term in order to use it in the context of an evolutionary explanation of the mind.

American students given the same question, because the latter were able to use the recognition heuristic (and American students' results were poorer than German students' when asked about cities in the United States). This research program has had important empirical results and applications: in medical diagnosis, for example, decision trees elaborated with two or three simple and quick questions usually lead, when well designed, of course, to better results than complex decision algorithms using much more information (Breiman, Friedman, Olshen and Stone, 1984, chapter 6). In these heuristic decision processes, no other information is needed, and the output is reliable (moreover, with engineered heuristics, sometimes the output is the best output possible, cf. Gigerenzer, 2007, chapter 9).

Carruthers argues that, instead of being encapsulated in the Fodorian sense, modules may just implement simple heuristics rules leading to fast and reliable outputs, and this is exactly what wide encapsulation captures. A given module can consult information from / in other modules, and this does not lead to informational explosion because it implements search heuristics that make the search frugal (just as Newell and Simon, 1976 argued). As Carruthers puts it: "[Evolution] will favor a *satisficing* strategy, rather than an optimal one [...] [and it] will favor a variety of search heuristics that are good enough without being exhaustive." (2006, 54)

Then again, this particular characterization of modules does not mean all modules have only the minimum to satisfy Carruthers' account. Some modules may actually have all the characteristics of Fodor-modules. But, and this is a point Carruthers emphasize, in the context of the massive modularity hypothesis, some of the characteristics attributed to these 'strong' modules must be removed in order to include all possible processes of the mind as being modular, even when they do not fit perfectly in the 'original' definition of 'module.'

In the end, Carruthers proposes his weakened notion of modules in order to integrate all cognitive processes and assign to them some of the characteristics of modules in a meaningful way. To use again the chemical analogy made earlier, it is as we had a number of nonmetals (cognitive processes), some of which we could identify as oxygen (Fodor-modules), others as arsenic (central modules, as we will see in 1.4), but we refused to

make the distinction between oxygen and arsenic in order to only have one notion to explain all of non-metal chemistry. His definition of massive modularity goes like this (this very long quote is necessary to capture the details given by Carruthers):

> Each of these systems will have a distinctive function, or set of functions; and each will have a distinct neural realization [...]. All of these systems will need to be frugal in their operations, hence being *encapsulated in the wide-scope sense, at least* [...]. Moreover, the processing that takes place within each of these systems *will generally be inaccessible elsewhere.* [...] Thus construed, the thesis of massive modularity doesn't require that the mind should be composed of systems that are encapsulated in the traditional narrow-scope sense (*although many might be*). Nor need all of these systems be domain specific in their output conditions (*although most are likely to be*). And while modules are function-specific, their algorithms needn't be [...]. In addition, while many modules will be significantly innate, or genetically channeled, *many will be constructed through some sort of (probably modular) learning process.* [...] I should stress, moreover, that there is nothing in these considerations to suggest that modules will be elegantly engineered atomic entities with simple and streamlined internal structures. (Carruthers, 2006, 62, my emphasis)

Yet, Samuels (2006) and Prinz (2006) are not satisfied with Carruthers' account, mostly because, in their view, its flexibility (I emphasized this flexibility in the quote above) renders the notion too weak to be controversial or philosophically interesting to any extent.

1.4 Samuels' account of massive modularity: weak central modularity and stronger modules

For Prinz (2006) and Samuels (2006), indeed, Carruthers' notion of module is so much weaker than Fodor's original thesis (1983), and also than most other accounts of modularity, that it does not say anything interesting or controversial about how the mind works. For Prinz, Carruthers' view is a mundane affirmation with very little theoretical consequences – for instance, it cannot help us refine theories or research paradigms. This criticism of Carruthers is similar to the one I made in 1.2 of Barrett's (2009) account of modularity. However, Samuels is more charitable, and he observes that

> [o]n some readings, the MM hypothesis is plausible but banal. On other readings, it is radical but wholly lacking in plausibility. And on still further (more moderate but still interesting) interpretations, it remains largely unsupported by the available arguments since there is little reason to suppose that *central systems* – such as those for reasoning and decision making – are modular in character. (Samuels, 2006, 37)

The first approach he focuses on is Barrett's reading of the massive modularity hypothesis. The second is a version of the strong central modularity thesis where all the modules are Fodor-modules[24], and the third identifies Carruthers' strong central modularity account with a weakened, but not wholly deflated, notion of module. This section will discuss a fourth reading of the massive modularity hypothesis, as developed by Samuels (2006).

Samuels (2006) defends a more traditional (closer to Fodor's) definition of module, but also advocates the adoption of a weak central modularity thesis. He suggests that the notion of Fodor-module might be adequate to understand many cognitive processes, but that we should also concede that some processes do not have the most important characteristics of Fodor-modules, such as domain specificity, encapsulation or cognitive penetrability.

He justifies his position by arguing that there is little support for massive modularity hypotheses advocated by the strong central modularity thesis, but that it would be an error to thereby conclude, as Prinz (2006) does, "that minds are not modular to any interesting degree" (Samuels, 2006, 52). There are, he believes, very good reasons to think that many peripheral processes are modular and, similarly, that it would be untenable to argue that it is not possible for any central processes to be modular, at least, to some degree. Regarding the extent of modularity, his solution is a middle ground between Carruthers (2006) and Prinz (2006), and this middle ground is, more or less, a return to a Fodor-like account of the mind, that is an approach we can identify with dual-process theories. For Samuels, Carruthers' account is unsupported by the current data and, according to what I developed in 1.3, it also has a problem similar to the one Samuels attributes to Barrett's account (2009).

---

[24] Evolutionary psychologists were believed to defend such an account (cf. Fodor, 2000), but they have since clarified where they are standing (cf. 1.2).

Then again, Prinz' complete rejection of modularity is not plausible (especially if we are to adopt a weakened notion of 'module').

Samuels believes there is "considerable evidence for relatively low-level modular mechanisms" (2006, 45), but that most data in favor of central modularity is interpreted as such in light of theoretical arguments in favor of massive modularity hypotheses he does not endorse (cf. Samuels, 2006, 42-45). Simpler, peripheral, processes might be able to explain the same data as does the central modules introduced by the advocates of massive modularity hypotheses. The clearest example is certainly the way the Wason selection task is used in order to argue in favor of a cheater detection module while there might be different explanations for the results obtained, such as perceptual biases (this is, of course, a much debated issue; cf. Clarke, 2004, chapter 4; Evans and Over, 2004, chapter 5; Houdé *et al.*, 2000; Stenning and van Lambalgen, 2008, chapter 3).

Moreover, some cognitive phenomena are difficult to explain without central (i.e. nonmodular) processes, making the case for a strong central modularity thesis very difficult. There is clear evidence we can combine and integrate concepts from different domains (defined in terms of function or in terms of content), use our background knowledge to make good inferences (inferences to the best explanation, abduction; cf. Fodor, 2000, chapter 3), there is also strong covariance between performance in different cognitive domains[25], and, finally, some disorders have an effect on all of central cognition (such as general mental retardation, e.g. Down's syndrome[26]), etc. There is also neural evidence for domain-general control mechanisms, consolidating Samuels' position

---

[25] Cf. Stanovich and West (2008, 686, table 8) for a list of the tasks and effects that correlate with cognitive abilities and those that do not.

[26] Samuels is aware that the existence of general mental retardation by itself does not prove the existence of nonmodular central systems. In spite of this, general mental retardation is most probably the least interesting of the points he makes. Still, it does not undermine his argument since what he claims is that "taken together, [the phenomena mentioned above] do strongly suggest the existence of nonmodular – domain general and unencapsulated – mechanisms for thought," (Samuels, 2006, 48) and these phenomena are compatible with weak central modularity hypotheses.

(Krug and Carter, 2010; Cohen and Lieberman, 2010). While these arguments do not guarantee that central cognition is not modular to *any* extent (versions of the weak central modularity thesis could still be plausible), the onus is on the 'strong central modularists' to explain these diverse observations about the alleged centrality of some cognitive processes. Samuels is quite harsh:

> [...] the prospects of accommodating the above phenomena without positing nonmodular mechanisms appear bleak; and in view of the lack of argument for MM, I'm inclined to think the effort of trying to do so is, in any case, wasted. (Samuels, 2006, 48)

While the solution is not to get rid of modularity altogether, Samuels' suggestion in the end is to adopt a middle ground, where there are clear instantiations of modules in the strong Fodorian sense (e.g. low-level perceptual modules) and nonmodular central processes (that might be modular to a small extent in many cases, in accord with the weak central modular thesis). Samuels is clear: "The situation is, in other words, much as Fodor advocated over two decades ago (Fodor, 1983)." (Samuels, 2006, 52)

1.5 Conclusion

In this chapter, I discussed the notion of 'module' and emphasized some of its controversial aspects. Strong central modularists argue for a version of massive modularity where the mind is *only* modular but, as I argued, they do not offer a satisfactory account as the notion of 'module' they use is either too weak and not very interesting, or too strong and not plausible. Some theorists however have suggested a very promising middle ground, between 'module nihilism' (*à la* Prinz, 2006) and strong central modularity: the weak central modularity hypothesis. By using a weakened notion of 'module' but accepting that there are also central processes that are *not* entirely modular, these theorists offer a more plausible account for the architecture of mind. This account is very similar to the approach identified by Clarke (2004) as the one defended by Evolutionary Psychologists (since, it seems they have abandoned such an account, cf. Ermer *et al.*, 2007).

What would such an account look like? Fodor, as I discussed quickly at the end of 1.1, thinks central cognition must be nonmodular: in this sense, he advocates one account of dual-process theories, where the two systems are input and output modules and central (or higher) cognition. While, as we will see in the next chapter, Stanovich, an important proponent of dual-process theories, seems to advocate such a position, he takes good care of elaborating a weaker notion of 'module' than the one described by Fodor (Stanovich, 2004, 37-44). It is indeed quite hard to develop an account where central cognition is not modular to *any* extent, and where peripheral cognition respects most of Fodor's criteria. Weak central modularity is promising in this respect: it has the important advantage of being more flexible than other accounts of modularity, especially the strong central modularity hypothesis. This hypothesis can also be seen as another label for some forms of dual-process theories, where there are both modular and nonmodular processes, divided along two systems (S1 is mostly modular, S2 is mostly nonmodular).

Samuels (2006) suggests a similar avenue and, in his more recent work (2009), he takes the idea of dual-process theories very seriously. However, there are still some problems with dual-process theories as they currently stand, and these problems will be the subject of the next chapter.

CHAPTER II

CAN THE MIND BE DIVIDED INTO (ONLY) TWO SYSTEMS? A CRITICAL OVERVIEW OF DUAL-PROCESS THEORIES:

In this chapter and the next, I will argue that there are at least two types of processes within the mind, one we can associate with Fodor-modules (even if Fodor-modules might not be the best existing characterization of these processes), and one we can associate with central processes (here, too, Fodor's characterization might not be the best). This difference is both functional and phenomenological (maybe neurological too), but there is more to it: there are, I will argue, two clusters of processes, each sharing a large set of characteristics, within the mind. These clusters however are not 'systems,' but there is still an important distinction to make between the two types of processes: in this sense, there are deep functional differences between them. In Chapter III, I will develop my account of dual-process theory in detail, but I will first, in this chapter, survey some of the existing dual-process accounts of the mind.

I will briefly overview what is meant by 'dual-process theories' by exploring ideas like Stanovich's (1999; 2004), but also by exploring how diverse are theories regrouped under this 'dual-process' label. Afterwards, I will explain and criticize Carruthers' (2006; 2009) account of dual-process theories. For him, there are no processes of the mind that cannot be characterized by what is labelled as 'System 1': the components of S1 are the modules of his massive modularity hypothesis. I will return to some key points of 1.4 and offer some evidence as to why we should doubt Carruthers' explanation of S2. Then, I will look at Lieberman's (Lieberman, 2007; 2009) version of dual-process theories. For Lieberman, there

are two systems to be distinguished: the reflexive system (X-system) and the reflective system (C-system). While interesting, I will show that Lieberman's account is inadequate as it is too rigid to explain many mental processes. I will finally suggest that, generally, there are important problems with how dual-process theories are currently advocated, and that it is necessary to adopt a different perspective, where the notion of system, which is too rigid to offer an adequate explanation, is abandoned (following Samuels, 2009). As mentioned, I will tentatively suggest one such account in Chapter III.

2.1 Brief overview of dual-process theories

Evolutionary psychologists[27] try to understand how the mind works in an evolutionary framework and their perspective has brought about many changes in how we conceive and think about the human mind. As explained in 0.2, understanding the mind as a set of processes able to work flawlessly and giving the right output each time cannot be right: the mind is a kluge, a kluge resulting from the processes of evolution by natural selection. Dual-process theories can account for this kluggish arrangement of processes (two systems having their own specificities and sometimes being in conflict) within the mind.

Table 2.1 gives an overview of the names attributed to each type of system in different dual-process theories and Table 2.2 gives an idea of the characteristics usually attributed to each.

---

[27] I am not referring to Evolutionary Psychologists, as I did in Chapter 1. I am talking here about psychologists working in an evolutionary perspective *in general* (cf. Buller, 2005; Marcus, 2008, 6-9; Stanovich, 2004, chapter 5) as I did in 0.2.

Table 2.1

Labels attached to dual-processes in the literature, aligned on the assumption of a generic dual-system theory. (From Evans, 2008, 257)

References	System 1	System 2
Fodor (1983, 200[0])	Input modules	Higher cognition
Schneider & Schiffrin (1977)	Automatic	Controlled
Epstein (1994), Epstein & Pacini (1999)	Experiential	Rational
Chaiken (1980), Chen & Chaiken (1999)	Heuristic	Systematic
Reber (1993), Evans & Over (1996)	Implicit/tacit	Explicit
Evans (1989, 2006)	Heuristic	Analytic
Sloman (1996), Smith & DeCoster (2000)	Associative	Rule based
Hammond (1996)	Intuitive	Analytic
Stanovich (1999, 2004)	System 1 (TASS)	System 2 (Analytic)
Nisbett *et al.* (2001)	Holistic	Analytic
Wilson (2002)	Adaptive unconscious	Conscious
Lieberman (2003) [and Marcus, 2008]	Reflexive	Reflective
Toates (2006)	Stimulus bound	Higher order
Strack & Deustch (2004)	Impulsive	Reflective

Table 2.2

Clusters of attributes associated with dual systems of thinking. (From Evans 2008, 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

Generic dual-process theorists typically use the labels System 1 (S1) and System 2 (S2) to identify two clusters of processes: S1 is understood as a group of many subsystems (cf. Stanovich, 2004) that function in parallel and are qualified as automatic, unconscious and fast while S2 is a domain-general system qualified with the 'opposite' characteristics. Of course, this brief categorization is heuristic: the distinction between each system is rarely made so clearly, and researchers do not agree on how to divide and distinguish each system (Evans, 2008; 2009).

There are too many accounts in the literature to discuss all of them (cf. Table 2.1). However, some are more salient or representative and some better illustrate 'extreme' versions of dual-process theories than others. Carruthers' (2006; 2009) and Lieberman's (2007; 2009) accounts are good examples of extreme versions, while Evans' (2008; 2009) and Stanovich' (1999; 2004; 2009) are probably the most influential.

The idea behind many dual-process theories is the following. For some evolutionary minded psychologists, as we saw in 1.2 and 1.3, the mind is massively modular, and each of these modules is a system that evolved in parallel with the others (cf. Carruthers, 2006, 12-28 for a detailed account of this 'argument from design'). Whether they accept the massive modularity hypothesis as advocated by Carruthers (2006) or Barrett (2009) or not (I do not, cf. 1.2, 1.3, 1.4 and 2.2), most researchers accept at least that there is an important set of modular processes within the mind. Understood in this wide sense, modular processes are systems using (mostly simple) heuristics working well in their ecological environment (Gigerenzer *et al.*, 1999) but that can sometimes produce less adapted responses, especially when applied out of their normal range, because of their inherent limits – from the standpoint of evolution by natural selection they do not need to be optimized: satisficing is enough. Our minds are environmentally bounded, viz. "the most important bounds that shaped our evolving rationality were not internal, mental factors, but rather external, environmental ones" (Todd, 2001, 52), and this idea is crucial if we are to understand how the mind actually works.

Some of these limits are studied by the heuristics and biases research program in psychology, and these heuristics psychologists offer a great deal of evidence in favor of

dual-process theories of cognition (Evans, 2003). In fact, according to Stanovich *et al.*, the tasks in this literature were specifically designed "to pit a heuristically triggered response against a normative response generated by the analytic system [S2 / Type 2 processes]" (Stanovich *et al.*, 2008, 254).

Kahneman and Tversky – the initial proponents of this research program – discovered that cognitive biases are not random: they follow a pattern because the mind relies on specific 'innate rules of thumb' (S1 / Type 1 processes), the heuristics Gigerenzer refers to, to produce its outputs. Thus, it is possible to design tasks in which these rules of thumb produce errors instead of their usual 'good response.' However, we also have deliberate and slow processes allowing us to find the right answer. This is, according to Stanovich *et al.* (2008), an important source of evidence in favor of dual-process accounts of the mind. Kahneman and Frederick agree and state:

> The 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 and Frederick, 2005, 267)

Dual-process theories like the ones referred to in the above paragraph abound in the literature, both in cognitive and social psychology. Evans (2008, 263) identifies three main currents: dual-process theories of reasoning, dual-process theories of judgment and decision making, and dual-process theories of social cognition. The first one is concerned with deductive reasoning, and tasks like the Wason card selection task and the belief bias paradigm[28]. The second includes, among others, the study of heuristics and biases (cf. Gilovich *et al.*, 2002) and that of the decision trees mentioned in 1.3. The third is one of the dominant paradigms in social cognition and is concerned with "the automatic and unconscious processing of social information" (Evans, 2008, 268) such as stereotypes and related attitude changes.

Here, I will be mostly interested in "attempts [...] to map various dual-process theories into a generic dual-system theory" (Evans, 2008, 256) rather than in a specific perspective on dual-process theories. One such attempt towards a generic dual-process theory is Stanovich's (1999; 2004; 2009).

System 1 and System 2 (or Subsystems 1 and 2; e.g. Leslie, 1994; Scott and Baillargeon, 2009) have various characteristics, and names, attributed to them, but the general idea remains basically the same (cf. Tables 2.1 and 2.2). As mentioned in Chapter I (1.1 and 1.5), Fodor (1983; 2000) is a dual-process theorist, even if he does not actively endorse the label, so is Carruthers (2006; 2009), hence the importance of their account of modularity – as it defines what their 'System 1' is. Many such accounts of the mind exist and they share at least two tenets, as Samuels (2009) explains. First, the distinctions made in Table 2.2 (or its variant a given author prefers) *align*, viz. "processes which exhibit one property from a column typically, though not invariably, possess the others" (Samuels, 2009, 131). For example, in Fodor's (1983) account, modules having one of the nine characteristics possess the others in most cases (but it could have only five out of nine). Second, the processes are taken to be part of either S1 or S2 – to use Fodor's account again, a process is either modular or part of central cognition. For now, I will explore Stanovich's account of dual-process theories as it is very influential in the literature, probably more so than Sloman's (1996; 2002).

For Stanovich (2004), S1 is in fact a great deal of systems that he identifies more or less with Fodor-modules, and which he calls collectively 'The Autonomous Set of Systems' (TASS). His TASS "refers to a (probably large) *set* of systems in the brain that operate autonomously in response to their own triggering stimuli, and are not under the control of the analytic processing system [S2]" (2004, 37) and many of its processes are modular. However, by modular he means something "less restrictive and therefore less

---

[28] Valid and invalid syllogisms with familiar or unfamiliar content. For example, if we present familiar content that makes sense to a subject, he is very likely to judge the syllogism valid even if it is not.

controversial than most conceptions of modularity in cognitive science" (Stanovich, 2004, 37). Basically, he rejects most of the characteristics of Fodor-modules and qualifies the systems of TASS as "fast, automatic and mandatory" (Stanovich, 2004, 40).

Stanovich's TASS processes are not necessarily innate – some of them can be acquired through learning[29] –, associated with fixed neural architecture or specific ontogenetic sequencing. His notions of encapsulation and impenetrability are defined more loosely (cf. Stanovich, 2004, 282). TASS processes work in parallel and some manipulate higher-level inputs and outputs[30] (but most do not). Finally, there is no conscious experience associated with TASS' operation (but its output *can* be conscious). Face recognition, theory of mind, fear, naïve physics, folk biology, child care, etc. are listed as 'cognitive modules' by Stanovich (2004, 44, table 2.2) and they are all part of what he considers to be TASS processes.

S2, for its part, has the converse characteristics: so, while parallelism, automaticity, domain specificity, low effort and unconscious characters of the processes are defining features of TASS / S1, S2 is serial and has "central executive control, conscious awareness, capacity-demanding operations, and domain generality in the information recruited to aid computation" (Stanovich, 2004, 44-45). Its most important features are certainly its central role in cognitive control and its ability to inhibit and override TASS' outputs with rule-based processes; in Stanovich's own words: "[...] the analytic system allows us to sustain the powerful context-free mechanisms of logical thought, inference, abstraction, planning, decision making, and cognitive control." (Stanovich, 2004, 47)

In a more recent version of his theory, Stanovich (2009) replaces S1 and S2 with Type 1 and Type 2 processes, divided among three 'minds,' three systems he identifies as the

---

[29] While Stanovich mentions this as a possibility (Stanovich, 2004, 38 & 42), he does not elaborate on his idea of a part of S2 acquiring autonomy, and it is not clear how it is made possible in the way he describes his S2 and in how he develops his dual-process theory.

[30] In other words, the inputs and outputs of TASS processes are not necessarily shallow.

autonomous mind (only Type 1 processes, it is the same as TASS), the algorithmic mind and he reflective mind[31]. The main interest of this distinction is to explain some experimental results (e.g. Stanovich and West, 2008) according to which S2 works in different ways depending on the task. However, Type 1 and Type 2 processes keep the characteristics Stanovich previously associated with S1 and S2. Examples of Type 1 processes are the TASS processes mentioned above, and S2 processes are divided between the algorithmic and the reflective minds. This difference between his two accounts does not imply much for Stanovich's architecture of the mind: the important thing to keep in mind for our purposes is that the processes identified are still grouped in systems. The difference is that they are grouped into three systems instead of two, because his two systems account was not handling some of his most recent research data very well (e.g. Stanovich and West, 2008; West, Toplak and Stanovich, 2008; cf. 2.4.1 for more details).

A parallel can be made here with the idea of the mind as a kluge from section 0.2: if the mind is a kluge, it is very improbable that the distinction between S1 and S2 can be made as neatly as most dual-process theorists pretend. I hope to show, in what follows in this chapter, that the assumption of such a neat division is misleading, and that the mind is probably not nicely divided in two types of processes.

---

[31] Stanovich (2009) divides the S2 of his previous account (2004) into two distinct systems: the algorithmic mind and the reflective mind.

The algorithmic mind refers to the information processing part of S2: Stanovich's examples are "input *coding* mechanisms, perceptual *registration* mechanisms, short- and long-term-memory *storage* systems [and access to them]" (Stanovich, 2009, 29, my emphasis). It explains *how* some tasks are processed, but not the reasons *why*. The reflective mind, answering the *why* part, contains the goals and beliefs associated with that goal. In Stanovich's words: "It is only at the level of the reflective mind that issues of rationality come into play. Importantly, the algorithmic mind can be evaluated in terms of efficiency [with IQ tests] but not rationality." (Stanovich, 2009, 30)

In his book, Stanovich mostly emphasizes what IQ test measure. For him, the algorithmic mind is responsible for the differences in general intelligence (IQ), and the reflective mind, for the difference in rational thinking dispositions (Stanovich, 2009, 33-34). This model helps him explain why the differences between cognitive abilities (as measured by IQ tests) and rationality (as measured by success to certain tasks) do not correlate for some tasks used to *test* the abilities associated with critical thinking and rationality (Stanovich and West, 2008).

This, I believe, points towards an important problem of dual-process theories as they currently stand. Before elaborating on these problems in 2.4, I will explore two 'extreme' versions of dual-process theories. First, Carruthers' account where there is no proper S2 to speak of; second, Lieberman's account where there are two systems having their own neuronal architecture (or, at least, their own neural activation patterns).

2.2 Carruthers and the multiple iterations of System 1 processes

As I described in 1.3, Carruthers (2006; 2009) argues that the mind is more than massively modular: it is *entirely* modular, viz. there is no cognitive structure in the mind other than modules (as he characterizes them). As a consequence, there is no deep functional difference, he believes, in the components realizing S1 or S2: the same components underlie both 'systems,' although different properties can emerge from their activity. S2, for Carruthers, is realized in multiple iterations, in cycles, of S1 activity, "rather than existing alongside the latter" (Carruthers, 2009, 112).

According to him, the distinction between S1 and S2 *does* exist at a functional and at a phenomenological level (but only at those two levels), where each system has different properties, the most distinctive being phenomenological. S1 is fast and unconscious and S2 is slower – since the cycles of S1 processing take time – and conscious, following the ideas of the global workspace theory suggested by Baars (1988; 1999), for who consciousness is the means by which modules are recruited in the global workspace that gives rise to the functional properties of S2. According to the global workspace account, inputs are processed by the modules that can do it, and their output is delivered to the global workspace, a kind of mental blackboard where outputs are made available to other modules (Poirier, in preparation). Once published, the outputs are processed by other modules that, in turn, process them as a new input before publishing a new output on the blackboard. This process continues until a solution is found or until the focus of attention changes (e.g. new inputs come in).

For Carruthers, action-rehearsal and utterances in inner speech (he insists on the importance of language) are what make possible the emergence of S2 out of cycles of S1 processing, and it explains why S2 processing is specific to humans[32] (as language is unique to humans). He believes that S2 "is realized in sequences of action-schema activations (often rehearsals of natural language utterances), with the sequences taking place (sometimes) in accordance with learned rules and inferential procedures" (Carruthers, 2006, 256-257). These multiple iterations of S1 processes are what give rise to conscious experience, thereby giving us the impression that two systems, or two types of processes, are at work. Multiple iterations are also what explain how conceptual content arises and, more importantly, why S2 operations are serial, as "only one action can be mentally rehearsed and globally broadcasted at a time" (Carruthers, 2009, 120).

Carruthers believes it is an advantage of his view that it removes "any need to regard System 1 and System 2 as distinct" (Carruthers, 2009, 120) in the deep functional sense. The distinction between S1 and S2 is replaced with his weak notion of module, encompassing all of the mind's processes, and he replaces S2 with the global workspace and multiple iterations of his modules. His account however has some shortcomings.

First, it is important to note that, as we saw in 1.3, Carruthers' notion of 'module' can be criticized as being *too weak*: he is not very strict about what counts as a module and he thus removes most of the explanatory advantages the notion could have in a framework where it is defined more precisely (cf. 1.3). Yet, his account of the S1 / S2 distinction can only work if one accepts this weakened notion – which I do not for reasons put forth previously. Nonetheless, I will argue below that, even if we accept his notion of module, his account of

---

[32] Action-rehearsal is described as possible for chimpanzees and *Homo ergaster* (or *erectus*) (Carruthers, 2009, 114), but Carruthers is not clear what makes S2 specifically human. Of course, inner speech is part of the answer but it only consists in "rehearsal of speech actions [that] gives rise to imagery" (Carruthers, 2009, 117). This is a problematic assumption made by Carruthers in his framework, but I do not have the space to discuss it in details here. Briefly: an important feature of the framework I will develop in Chapter 3 will allow me to consider the 'linked to language' line of Table 2.2 in a broader framework where being linked to language is only one characteristic among many others (cf. footnote 44, p. 63).

the S1 / S2 distinction encounters two major difficulties: 1) explaining executive control as well as 2) accounting for various observations in neuroscience that suggest the existence of central / nonmodular processes. It also appears that his most original claims are predictions, none of which are currently confirmed (see below; also, Samuels, 2006 makes a similar claim).

About this last remark: Carruthers specifies that adopting his account of S2 as action-based predicts (i) that "System 2 thinking skills should be acquirable by imitation and instruction, and that sequences of System 2 reasoning should be shaped by belief about the ways in which one *should* reason" (Carruthers, 2009, 121). Plus, (ii) he expects this will account for variation of S2 across cultures, and (iii) explain why failures or problems with S1 (such as in subjects with particular diseases or with brain damage) might impair S2. Finally, (iv) he also believes his account is able to resolve problems of current dual-process theories.

While the first prediction is original, it is currently not confirmed, and some evidence suggests it will not lead to an adequate model of how we reason and acquire new thinking skills (including but not limited to critical thinking; see below my discussion of Houdé *et al.*'s results). In addition, his second and third predictions are certainly not unique to his account. For example, Stanovich (2009, chapter 10) offers interesting explanations of cultural variations with his account of reasoning and of the use of mindwares (Stanovich and West, 2008). Moreover, *any* dual-process account has something to say about how a defect of S1 might impair S2: for example, children with attention deficit disorders will have trouble concentrating, thus having difficulties sustaining the attention required to override certain S1 processes. This impairment of their performance will have a direct effect on their competence. Certain S1 processes are necessary for the execution and realization of some S2 processes; in other words, there are many overlaps between processes of both systems and any dual-process theorist recognizes this almost trivial fact.

Lastly, he does not argue convincingly that he is addressing actual problems dual-process theorists might encounter. He rather highlights the lack of details on some important issues. For instance, Carruthers briefly argues that the relationship between S1

and S2 is not clear: the override mechanisms are S2 processes interacting with S1's. Indeed, the interaction between both systems is rarely discussed, but many accounts, especially Stanovich's, can offer a convincing explanation of these relations[33]. The problems Carruthers identifies are mostly due to the lack of research in an emerging field; they do not at the present appear to be unsolvable problems of dual-process theories.

Let's return to the two problems that Carruthers's account of the S1 / S2 distinction faces. The most acute is that it is not clear how executive control, a crucial aspect of S2 processing, emerges out of the automatic processes of S1, nor how executive control can override / intervene on S1 processing. He identifies the functional and phenomenological differences between S1 and S2, but, since he rejects any deep functional difference between them, he does not have the tools to explain how the iterations of S1 processes can produce new functions such as executive control. He can, and does quite well, explain the specific character of S2 processes, as they are the result of the multiple iterations of S1 (hence, S2 is slower and serial), but his account lacks the capacity of explaining *new capacities*, viz. how new properties can emerge from the multiple iterations of modules. My point is that, although it is very likely that Carruthers' account can explain many cognitive phenomena, I doubt that it can explain *all* of them.

Moreover, Carruthers – in his model (Carruthers, 2006; 2009) – does not have the tools to explain how new brain regions become activated with a modification of the instructions given to solve a problem without any *ad hoc* hypothesis. While one could say that new brain areas have to be activated for the reiteration to occur, or that they become activated at a certain point of the iterations (see below for what this suggestion would entail), this would be *ad hoc* as it is not considered; there are no 'new' processes intervening in his account. This is

---

[33] As mentioned, Stanovich offers a complete and detailed account of these interactions while discussing the acquisition and use of mindwares (Stanovich, 2009; Stanovich and West, 2008). Also, Stanovich never rejected that impaired parts of TASS or impaired Type 1 processes could have effects on S2 / Type 2 processes. Both systems are independent of each other, but S2 uses S1 inputs: for example, it is not possible to reason about *any* task if we cannot perceive it, understand it (language recognition), concentrate on it, remember how to proceed, and so on.

what Houdé and Moutier's (1996; 1999) work in neuroscience illustrates. It suggests that the brain areas that are activated before and after the subjects receive a training in logic to solve Wason selection task-type problems are the same, and they rarely give the right answer. However, in a more recent study (Houdé *et al.*, 2000), Houdé and his colleagues showed that subjects trained to inhibit their perceptual bias (cf. Evans, 1998; 2003), viz. made aware of their cognitive conflict, were better at solving a selection task *and* the brain areas activated were not the same anymore (the neural activation patterns shifted from the occipital lobe to the frontal lobe). As the authors explain:

> The most striking result obtained here is the change in the cortical anatomy of reasoning, which shifted from the posterior part of the brain on the pretest to a left-prefrontal network on the posttest, thereby reflecting the change in the subjects' reasoning strategies. (Houdé *et al.*, 2000, 723)

How can Carruthers explain this shift if the subjects are only iterating S1 processes? It seems hard to explain the inhibition of the subjects' initial answer while only learning and repeating the task does not work (Houdé and Moutier, 1996; 1999). Here is how he could perhaps explain it: the erroneous S2 responses result from the activation of S1 processes A, B and C, which each corresponds to a specific brain activation pattern. In Houdé and Moutier's (1996; 1999) studies, the subject learns to explicitly use the process D involving an explicit rule, viz. linked to language, but the processing made by A, B, C and D does not change the subject's answer because the processes that caused the errors in the first place are still executed. And since process D is of the same type as A, B, and C, the same general brain regions are activated. When the subject is then made aware of his perceptual bias, he learns something that makes him use process E, through an explicit instruction of some sort, and then succeeds at the task (Houdé *et al.*, 2000) either because E has the effect of inhibiting C, or because E is such as it overrides C. Hence, the processing of A, B and E allows the production of the correct answer. And since E involves the inhibition of activity that is located in a new part of the brain, this part lights up on the scan[34].

---

[34] One frequent critique of current imaging techniques is that they cannot distinguish between activation-in-the-course-of-processing and activation-in-the-course-of-inhibiting-processes.

This explanation fits Carruthers' framework perfectly as action-rehearsal through inner speech could very well account for this change as it would activate a new process or inhibit the one producing the error (as described in the last paragraph), potentially showing differences in the neural activation patterns – imagining an image will activate visual areas, while activating the memory of smells or sounds will activate brain areas corresponding to that sense.

But this calls attention to the previous difficulty arising from Carruthers' account, as it does not give an explanation of what determines the controlled execution of inner speech driven action-rehearsal in one way rather than the other, viz. why one set of instruction will be preferred, in a given context, to the other one. In Lieberman's framework, this is where self-control would be involved. In the situation described by Houdé *et al.*, conflict detection and the capacity to override when needed are partially explained by Carruthers' framework but the crucial aspect of executive control is not. In fact, what happens in Houdé *et al.*'s experiment is easy to account for since the instruction that is needed to solve successfully the problem is given by the experimenter, or is made prominent by the experimental set-up. The experimenter acts just like the executive control would, by specifying the correct set of instruction to use in a given situation. But what about a situation where two or more sets of instruction are available to the subject: how is the choice made, and why? The controlled aspect of these processes eludes Carruthers' model: how is the correct or incorrect instruction activated, and then chosen? How does the instruction gets involved in this particular iteration, why is it remembered and what is our control on when and why it is activated? It is however important to be careful in answering these questions: as Stanovich (2004, 44-47) remarked, it is essential to avoid the homunculus problem.

Results similar to Houdé and colleagues' (2000) are well accounted for in Lieberman's framework (Lieberman, 2007; Satpute and Lieberman, 2006), and it might help us understand

what exactly goes on in the brain in such situations[35]. There needs to be some sort of executive control intervening. Specific neural activation patterns related to self-control have been identified: it is the alleged function of the right ventrolateral prefrontal cortex (Cohen and Lieberman, 2010). More generally, this difference is correlated with different neural activation patterns depending on the task at hand, sometimes associated with Type 1 processing, and sometimes with Type 2 (Goel, 2007).

2.3 Lieberman on neurological and deep functional differences

Lieberman's (2007; 2009; Satpute and Lieberman, 2006) position is in opposition to Carruthers' as he argues that there are important differences, *deep functional differences* as I would put it, between what he calls the reflexive system (X-system, or S1) and the reflective system (C-system, or S2). Not only are the cognitive processes he singles out associated with different areas of the brain, but he associates both of them with different characteristics, from the phenomenological to the representational and evolutionary (cf. Table 2.3).

I will not go into the details of each of the characteristics enumerated in Table 2.3, as they are pretty much self explanatory, but it is interesting to see how Lieberman's account overlaps with dual-process theories in general, that he also adds some important characteristics, mostly coming from his field of inquiry, social cognitive neurosciences. His view introduces clear distinctions, as each system is associated with precise characteristics. Lieberman has the tools to make, with his account of dual-process theories, many

[35] Stanovich's framework (Stanovich and West, 2008) can account for this shift in neuronal activation too. In fact, Stanovich's account is really close to Lieberman's, as he agrees there are domain general processes that can account for much of higher cognition.

predictions[36] – for example, he can infer, from the activation of a given brain region (cf. Figure 2.1), the phenomenological and representational characteristics, and the converse is also true.

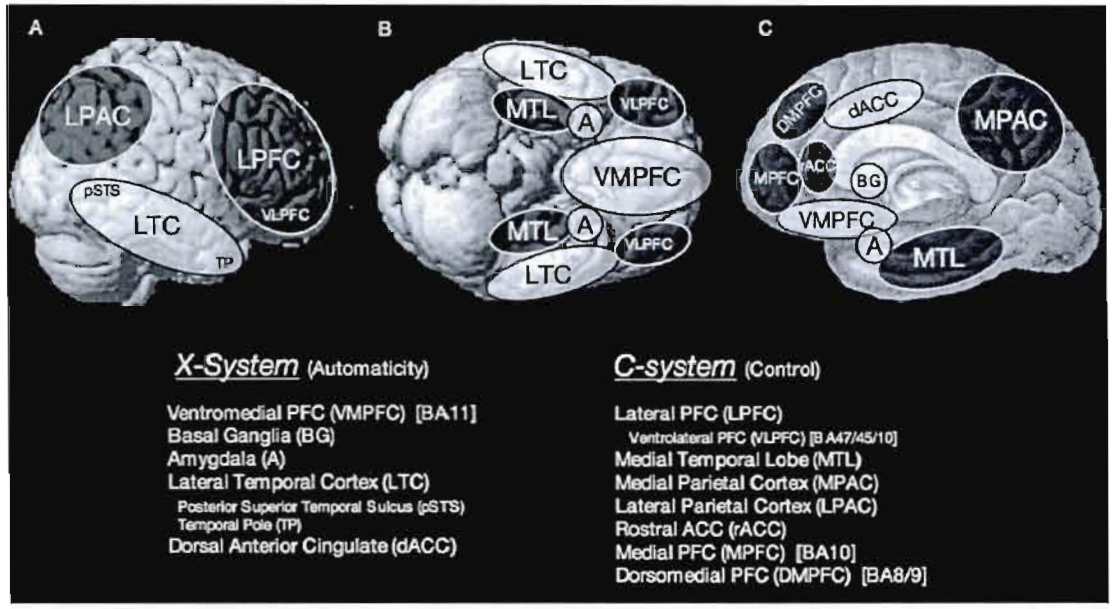

Figure 2.1 Hypothesized neural correlates of the C-system supporting reflective social cognition (analogous to controlled processing) and the X-system supporting reflexive social cognition (analogous to automatic processing) displayed on a canonical brain rendering from (*A*) lateral, (*B*) ventral, and (*C*) medial views. (From Lieberman, 2007, 262.)

---

[36] Lieberman (2009) suggests this difference in how is the brain activated when reflective and non-reflective processes are involved gives strong evidence against the psychological zombie hypotheses, but this is not important in the current discussion. I am mostly interested in how he characterizes each system in order to develop his argument.

Psychological zombie hypotheses differ from Kirk's (1974) zombie hypothesis: a psychological zombie hypothesis "suggests that our behaviors and judgments are produced by an 'inner-zombie' whose mental work does not depend on conscious awareness and that those mental operations that are typically accompanied by conscious awareness do not rely on awareness to generate the operations and their outputs" (Lieberman, 2009, 293). This kind of hypothesis suggests that awareness is superfluous. Thought experiments around blindsight tend to go in this general direction (Holt, 1999), but Lieberman argues, successfully I think, that this line of argument is faulty.

Table 2.3
Characteristics of the X-system and C-system. (From Lieberman, 2009, 294)

	X-system	C-system
Phenomenological characteristics	Non-reflective consciousness Feels spontaneous or intuitive Outputs experienced as reality	Reflective consciousness Feels intentional and deliberative Outputs experienced as self-generated
Processing characteristics	Parallel processing Fast operating Slow learning Implicit learning of associations Pattern matching and pattern completion	Serial processing Slow operating Fast learning Explicit rules of learning Symbolic logic and propositional
Representational characteristics	Typically sensory Representation of symmetric relations Representation of common cases Representations are not tagged for time, place, ownership, identity	Typically linguistic Representation of asymmetric and conditional relations Representation of special cases (e.g. exceptions) Representation of abstract features that distinguish (e.g. negation, time, ownership, identity)
Evolutionary characteristics	Phylogenetically older Similar across species	Phylogenetically newer Different in primates or humans
Moderator effects	Sensitive to subliminal presentations Relation to behavior unaffected by cognitive load Facilitated by high arousal	Insensitive to subliminal presentations Relation to behavior altered by cognitive load Impaired by high arousal
Brain regions	Amygdala, ventral striatum, ventromedial prefrontal cortex, dorsal anterior cingulate cortex, lateral temporal cortex	Lateral prefrontal cortex, medial prefrontal cortex, lateral prefrontal cortex, medial posterior parietal cortex, rostral anterior cingulate cortex, medial temporal lobe

The two systems usually are in harmony in Lieberman's view, even if they are sometimes competing one with the other – especially in experimental settings. More importantly however, they are *separable*, something impossible on Carruthers' view as he believes S2 is realized in multiple iterations of S1 processes.

In most of the literature, the systems even seem to be *always* in competition with one another. It is important to note that in the experimental tasks, the goal *is to differentiate* the two systems, to identify their distinguishing features. This is why X and C systems appear as independent. But, out of the lab, "in everyday life, most tasks probably rely on both systems simultaneously" (Lieberman, 2009, 313) and X and C systems are such that they work together to produce reliable outputs. In Lieberman's view however, this collaboration between systems remains this: two distinct systems, systems we can identify at various levels including the neurological, collaborating.

Lieberman goes as far as to say that the systems are operating "upon different principles" (2009, 313) and he relies on various neuroimaging studies to illustrate and justify his claim that different neurological mechanisms are at work. The study by Houdé *et al.* (2000) discussed above shows a similar phenomenon happening when subjects change their reasoning strategies in a particular reasoning task. In Figure 2.1, from Lieberman (2007, 262), the brain regions associated with each system are illustrated.

There are, as Lieberman emphasizes, specific neural activation patterns depending on whether the tasks involve controlled processes or not (cf. Goel, 2007). He argues this difference suggests there is a "core-processing distinction" (Lieberman, 2007, 276), where there is a strong link between the neural activation patterns and the phenomenon observed. According to Evans, "this research program provides perhaps the strongest basis in the literature for maintaining some form of *dual-system distinction*" (2008, 270, my emphasis).

However if, as I argued in section 0.2, the mind is a kluge, it would be very surprising to discover that it has indeed two perfectly distinct systems, collaborating to "achieve the best outcomes" (Lieberman, 2009, 313). Just as the outcomes of evolutionary processes are usually inelegant, we can, I believe, suppose the same is true with the human mind and, as *it is* the result of an evolutionary process, Lieberman's 'panglossian' account is not very plausible: we cannot divide the mind so neatly.

As we will see in the next section, the characteristics attributed to both systems by most dual-process theorists, and Lieberman is particularly vulnerable to this critique, are indeed too rigid to identify all the complex and rich processing the brain is capable of. While

Lieberman's evidence is compelling (he even identifies convincingly some domain general processes that seem distinct from the X-system processing; cf. Cohen and Lieberman, 2010), his account is probably not subtle enough to characterize correctly some of the cognitive processes at play. He does identify very important and interesting characteristics of *some* processes, but his framework has to be, at least, weakened in some respects.

2.4 Unsystematic systems

As I will now argue, the most important problem that accounts of dual-process theories relying on the characterization of systems face is that the lists of characteristics attributed to each system are too rigid. Evans, even if he agrees that Lieberman's evidence is compelling, mentions that a "close inspection of the evidence suggests that generic dual-system theory is currently oversimplified and misleading" (Evans, 2008, 270). He makes this point because it is not possible to link in a coherent way all the characteristics enumerated in Table 2.2 (or those in Table 2.3), but also because some accounts of dual-process theories are not fully compatible with each other (Evans, 2009) – which partly explains why it is difficult to construct a generic account, as we examined in 2.1. Those differences make it impossible to map all the processes of dual-process theories in a generic framework such as those currently advocated.

While building generic dual-process theory accounts, to explain these differences between dual-process accounts, but also to suggest a theory where the relevant data is accounted for, some authors have begun to suggest amendments to one of the original frameworks. While some are suggesting we should distinguish between up to four types of processing in a multiple-systems framework (Moshman, 2000), or even more complex architectures, others are arguing against each of the proposed criteria to distinguish the systems, using examples and counter-examples to show the account in question is too limited or has an important problem when trying to explain some of the data, sometimes ultimately rejecting dual-process theories altogether (e.g. Osman, 2004). One of the important and influential change introduced in the recent literature is Samuels' (2009) proposal of talking of

'Types' of processes instead of 'Systems' (I will return to this argument at the beginning of Chapter III), such as it is usually the case with the S1 / S2 distinction. It allows more flexibility to the accounts of dual-process theories proposed, but there are still some problems to face.

Of course, some of the dual-process accounts, such as Fodor's (1983), are obviously problematic as they offer very limited account of how the mind works; Fodor's modularity only identify a small subset of cognitive processes, peripheral cognition and the input modules, and central cognition, only identified with a vague definition. As I argued in 1.4, an account resembling Samuels' (2006) characterization of weak central modularity hypothesis is surely more plausible and has more potential to yield fruitful research programs. However, even moderate accounts encounter difficulties, some that might be insurmountable. Here are two examples: Stanovich's (2004; 2009) and Evans' (2009) accounts.

2.4.1 Stanovich's three minds

In 2.1, I suggested that Stanovich's (2004) framework has some problems, bringing him adopt a modified account of his original position in his 2009 book. As a reminder, Stanovich replaced the distinction between systems S1 and S2 with a distinction between Type 1 and Type 2 processes, these types of processes being divided among three systems, the autonomous, the algorithmic and the reflective minds.

Stanovich describes S2 as "more strongly associated with individual differences in cognitive in computational capacity (indirectly indicated by tests of intelligence and cognitive abilities – and more indirectly tapped by indicators of working memory)" (2004, 36). However, Stanovich and West (2008) encountered some experimental results incompatible with this framework: they show that for *some* tasks involving cognitive control, performance is *not* correlated *at all* (or very poorly) with cognitive ability (as measured by IQ tests). For

example, the myside biases[37] (Stanovich and West, 2007) or the anchoring effects[38] do not correlate with cognitive ability, while belief biases sometimes do (Stanovich and West, 2008, experiment 8). In other words, success in some cognitive tasks is correlated with cognitive ability (IQ), while it is not correlated with cognitive abilities in other tasks, something hard to account for with a 'single, unified, S2' as Stanovich's (2004) S2.

Separating the processes between two types, distributed among three minds, allows him to explain this problem in a convincing framework. The new framework also provides useful tools and an interesting explanation to what is going on in the study by Houdé *et al.* (2000). Moreover, the introduction of the distinction between the algorithmic and the reflective minds, where there was only a reflective mind before, allows Stanovich (2009) to understand better these differences in the correlations between success and cognitive ability, but the processes involved are more or less similar to the processes he described as comprising S2 in his 2004 account; they only are distributed differently. Still, there are processes hard to classify by Stanovich's most recent account, such as preattentive processes, since he is mostly interested in reasoning and reasoning strategies and not, for example, in the 'consciousness' axis of the distinction between S1 and S2.

---

[37] 'Myside biases,' also called confirmation biases, refer to "the tendency to evaluate propositions from within one's own perspective when given no instructions or cues (such as within-participants conditions) to avoid doing so" (Stanovich and West, 2007, abstract). Interestingly, Mercier and Sperber (forthcoming) argue this is a *feature* of reasoning because it is useful in the construction of arguments.

[38] The 'anchoring effects' are a set of cognitive bias observed when people make an estimate from an initial value that they must adjust: people adjust the value to a certain extent but the adjustment is insufficient. The result is that "different starting points yield different estimates, which are biased toward the initial values" (Tversky and Kahneman, 1974, 1128). In other words, when asking people to estimate a value they do not know, if we give them a small number, they will tend to make a smaller estimate than if we give them a bigger number.

2.4.2 Evans' Type 3 processes

Evans (2009[39]), encountering a different problem than Stanovich's, suggests we should adopt, just as Stanovich, Type 1 and Type 2 processes, but that we should moreover add Type 3 processes to explain some characteristics of the mind that are impossible to account for with only Type 1 and 2. Type 3 processes refer to the processes allowing "resource allocation, conflict resolution, and ultimate control of behavior" (Evans, 2009, 48). Indeed: in an account where the distinction between the two systems, or two types of processes, is rigid and where S1 processes are fast, implicit and demand low effort and where S2 processes are controlled and conscious, it is difficult to explain how to classify pre-attentive processes, such as the control of attention and the detection of cognitive conflict[40]. Type 3 processes administer the interactions between Type 1 and 2 processes. In the end, these Type 3 processes are somewhat similar to Baars' account of attention control in his general workspace as advocated by Carruthers (2006; 2009).

Evans' motivations are linked to "recent evidence [suggesting] that the mind does detect conflict, even when we are not conscious that it is doing so" (Evans, 2009, 49). The example given by Evans is that of a driver having a conversation while 'unconsciously' driving: if a hazardous situation presents itself – a car suddenly braking, a moose coming out of the

---

[39] It might seem strange that Evans (2009) advocates the view I will describe below *after* the publication of his *Annual Review of Psychology* article, where he suggests we use the distinction Type 1 / Type 2, but the article published in 2009 is part of the proceedings of a 2006 conference. For the same reason, Evans (2008) mentions Samuels' (2009) argument, before the publication of the article (he quotes the conference presentation in the paper).

[40] Usually, in dual-process theories, the allocation of resource and the resolution of conflict are attributed to S2 (cf. Cohen and Lieberman, 2010) since its role is to monitor S1 activity and decide whether to intervene (Evans, 2009, 48), but it is an unconscious process – something incompatible with S2 or Type 2 processes' characterization. But we cannot characterize these pre-attentive processes as Type 1 since the role of these processes is to override default processes – a characteristic attributed to S2 or Type 2 processes. Evans (2007) solves this problem by introducing a third parameter describing "the probability that a type 2 rather than type 1 process will take control [of the behavior]" (Evans, 2009, 47), which he explains by using Type 3 processes in his 2009 article (for more details, cf. Evans, 2009, 46-50).

woods, etc. – attention *automatically* shifts. This is, according to Evans, the doing of Type 3 processes (or System 3). Such a process is neither Type 1 nor Type 2, as it has characteristics from both types of processes: compared to a rigid '(only) two types of processes' framework, Evans' alternative might be a reasonable one.

While the introduction of a new type of process is problematic, at least to a certain degree, for Lieberman's account, the idea I want to emphasize here is the difficulty of any account that has two clusters of cognitive processes (whatever you call them) identified each with a precise set of characteristics. Such accounts inevitably face difficulties when they try to explain processes that are not easily identifiable with these two systems, thus S1 and S2 because of their rigidity are ill-conceived. As it is, there always will be a process or a set of processes that is hard to understand in any two-track model – emotions might just be in this category (Darlow and Sloman, 2010; de Sousa, 2010). I believe that my account, inspired from Samuels' (2009) critique of dual-process theories, has the potential of offering a more comprehensive account, a framework that I will develop in Chapter III.

2.5 Conclusion

In this chapter, I discussed various accounts of dual-process theories and argued that the current accounts are on shaky grounds. I showed that some of the representative theories currently available encounter problems when explaining some phenomena. The 'Systems' or the 'Types' are inferred from a very small set of data, for example the set of data coming from the psychology of reasoning (dual-process theories of reasoning originate from the analysis made by Wason and Evans' in their 1975 article), and then the principles found are applied to larger sets of data. Facing complex sets of data and nonstandard processes, these theories must be adapted – sometimes by adding a new set of processes along the way – and these epicycles 'complexify' the theory. As illustrated in 2.4.1 and 2.4.2, many of these 'epicycles' – not always compatible between them – are added to make sense of the data in a given dual-process framework, and they are, in this sense, *ad hoc*.

Stanovich's and Evans' theories encounter difficulties that are representative of those found in this literature: the proposed description of each system, or of each type of processes, gives rise to two restrictive categories in which some cognitive processes are hard – and sometimes impossible! – to map. Lieberman's account is the clearest example: his framework is very interesting and would allow for much explanatory power, but his distinction between the C-system and the X-system can only explain a limited number of cognitive processes – it is not a theory able to explain *all* cognitive processes.

As for Carruthers atypical account, it helps resolving some problems – and it is quite plausible that many Type 2 processes may in fact the result from (only, or almost only) iterations of Type 1 processes. I think however the burden of proof is in Carruthers' hands: he would have to show that everything we want to account for can be explained in his framework, especially executive control. These problems are not alien to those he encounters by using his too weak account of 'module.' Because he wants to explain everything cognitive with a single notion, Carruthers' very ambitious program quickly becomes unviable.

In my last chapter, I want to suggest an account of dual-process theories that solves many of the problems identified in the first two chapters of this dissertation, but first I will examine Samuels (2009) suggestion of replacing the division of the mind in two 'Systems' by a distinction between 'Types of processes,' viz. he suggests the idea that there is no plausible way to argue there are a single S1 and a single S2.

CHAPTER III

CONCEPTUAL SPACE AND THE TYPE 1 / TYPE 2 DISTINCTION AS A HEURISTIC: ANOTHER ACCOUNT OF DUAL-PROCESS THEORIES

As I hinted at when I discussed Stanovich's (2009) and Evans' (2009) recent accounts, a current trend in dual-process theories is to abandon the division into 'Systems,' in order to prefer the distinction between 'Types of processes.' This is an interesting and important development as it gives better tools to consider the large variety of cognitive processes. Still, the 'Types' accounts as they are currently discussed in the literature are not exempt of problems, as I suggested in 2.4.

A first problem for dual-process theories reconceived as a distinction between 'Types of processes' is that abandoning the 'Systems' accounts makes dual-process theories lose some of their explanatory power: an account such as Lieberman's (2007; 2009) is really useful when we try to understand how the mind might work; as Samuels writes: "Dual-process theorizing is worthy of serious consideration because it earns its *explanatory keep*." (2009, 138) Yet, it is clear that if there were no identifiable 'System' whatsoever, then it would be preferable to abandon this account for a more adequate one, and I argued that, even if Lieberman's account is interesting to account for important aspects of cognition, it is too rigid to explain *all* of it. Moreover, as I will show in section 3.3.3, it is possible to get similar explanatory power with another conception of dual-process theories. The second problem is that, with the rigidity of the 'Types' proposed in current accounts (Evans, 2009; Samuels, 2009; Stanovich, 2009), we encounter similar problems as those with the 'Systems'

account. Evans' (2009) introduction of Type 3 processes is a clear example of this kind of difficulty.

There are two central goals in this chapter: first, to show that dual-process theories as currently conceived are untenable – whether we characterize this 'dual' distinction as between 'Systems' or 'Types of processes,' and, second, to lay the foundations of a new framework. In this new framework, much of the explanatory power of current dual-process theories is preserved, the labels 'Type 1' and 'Type 2' acquire a new meaning, and does so without cutting corners or oversimplifying the complexity of the many cognitive processes under scrutiny.

I will begin with a presentation of Samuels' (2009) account of dual-process theories, where he argues that the mind has two 'Types of processes.' I will mostly center my analysis on the problems he identifies with the traditional dual-process theories (i.e. the theories dividing the mind into two 'Systems'). I will then show there are some problems with his own account, impossible to solve in his framework; in a nutshell, his 'Types' are still too narrowly defined to explain *all* of cognition.

At the end of 3.1, I will show that using the idea of a continuum to analyze these 'Types of processes' might help resolve some of Samuels' difficulties. In 3.2, however, I will suggest that even this idea of a continuum (cf. Hammond, 1996) has problems and I will argue for a framework that has the potential of solving much, if not all, of the problems of current dual-process theories.

Briefly, I will argue that by considering the distinction between 'Types of processes' in terms of continua rather than in terms of an either-or distinction, we will be able to make a first step, resolving some of Evans' (2009) concerns. However, a single continuum will not be enough to preserve each process' distinctive characteristics: many continua forming a n-dimensional conceptual space (where n is the number of characteristics considered) should provide us with a more adequate framework. After arguing for this framework in 3.2, I will finally discuss three of its advantages in 3.3.

3.1 From 'Systems' to 'Types': going beyond

The idea of abandoning the 'Systems' accounts to adopt a division between types of processes comes from Samuels (2009). What Samuels suggests is that we should no longer talk of 'System 1' and 'System 2' in order to talk in terms of 'Types of processes.' The 'Systems' accounts refer to two tokens (S1 and S2), while, in the 'Types' account he proposes, these labels identify types instead of tokens, viz. "it seems likely that there are both many system 1s and many systems 2s." (Samuels, 2009, 138). He argues that, on the one hand, many dual-process theorists agree that S1 is a set of Type 1 processes (e.g. Stanovich's TASS) and that, on the other, there is no solid evidence that S2 is unified in any way (e.g. Stanovich's algorithmic and reflective minds; cf. Samuels, 2009, 137, Consideration 2). Different S2 accounts attribute different properties to S2: Samuels identifies functional differences (planning is different from deductive inference[41]) and there are also differences in computational demand for various (S2) processes. Samuels' point is "that the most plausible view of this sort is one that reconstructs the original S1 / S2 distinction [what he calls the 'Token Thesis'] as a distinction between kinds or *types* of psychological systems" (Samuels, 2009, 145), viz. what he dubs the 'Type Thesis.'

While he suggests we talk in terms of 'Types of processes' rather than in terms of 'Systems,' Samuels remains vague about how we should identify these types of processes and about what these types identify. With the 'Type Thesis' however, Samuels still maintains that characteristics from Table 2.1 co-vary, but he is no longer committed to the 'Token Thesis' where there are a single S1 and a single S2.

---

[41] "[...] planning is centrally concerned with the guidance of action – with identifying sequences of behaviors that collectively facilitate the attainment of our goals. Moreover (and presumably because of this) planning involves a quite different mapping from inputs to outputs than those found in deductive inference or causal reasoning. Most obviously, a planning process takes both beliefs and goals (or desires) as input and generates plans (or intentions) as output, whereas other sorts of process – deductive inference or causal explanation, for example – do not." (Samuels, 2009, 137)

In a way, a problem with the 'Type Thesis' is the same as the one researchers encounter with the 'Systems' accounts: the processes examined are rich, complex and hard to classify in fixed categories. This problem is less troublesome with the 'Types' accounts, where it is possible to attribute some characteristics to a process, but not others, but difficulties are not entirely eliminated. Samuels identifies three challenges for the 'Types' account, challenges that illustrate very well some, but not all, of the difficulties I am referring to: the specification problem, the crossover problem and the unity problem. What I claim is that the crossover problem cannot be resolved in a satisfying way with Samuels' thesis.

The specification problem might be the most complicated of the three, and the one necessitating the most research: how should we specify each type of process? Which characteristics should be retained? Characteristic clusters attributed to S1 and S2 are often criticized because, in the 'Systems' framework, it is easy to find a process that is at odds with the other processes of a given system (or with the systems themselves). (Re)Conceiving dual-process theories as identifying two types of processes instead of two systems (tokens) will partly resolve this difficulty, and it allows researchers to revise the characteristics attributed to each process. Even then, it is still not easy to identify which characteristics are the most important ones.

Samuels believes the specification problem will be resolved when the current hypotheses regarding the attribution of characteristics to processes is more refined; only more empirical research can help achieve this. While doing so, it is necessary not to specify each type too narrowly, as we would encounter problems such as the ones I highlighted in 2.4, especially when I exposed some of the problems with Evans' (2009) position.

Samuels' 'Types' account hints at a possible solution to the second problem, the crossover problem. The crossover problem might be the most important problem of Lieberman's (2007; 2009) account and, one could argue it is also a problem of all generic dual-process theories: Samuels explains the crossover as the possibility of a process to have characteristics from both columns of Table 2.2, since "the characteristics exhibited by cognitive processes are not amenable to a clean bipartite division into two property clusters" (Samuels, 2009, 140). There are some processes hard to categorize in either S1 or S2 as they

exhibit properties from both. The crossover problem appears in the description of many processes, but Samuels suggests it might be caused by some characteristics that should not have been used to distinguish the two types of processes, such as evolutionary recency, rather than by many processes exhibiting characteristics proper to different types. He gives, as an example, judgments of numerical magnitudes, which exhibit S1 and S2 properties at the same time: it is evolutionarily recent like a (Type 2) characteristic, since the number Stroop effect[42] is possible only with known numerals – Arabic or otherwise (which are learned through Type 2 processes) –, but it is nevertheless fast, automatic and demands low effort (all Type 1 characteristics). In the end, counterexamples such as this one, these 'hard to characterize' processes impose, Samuels believes, "modest revisions" that provide "no serious grounds for rejecting the Type Thesis as such," as long as these revisions are not "too numerous or too extreme" (Samuels, 2009, 140-141). In other words, there is no real crossover problem for Samuels: only characteristics that should not be incorporated in the characterization of either Type 1 or Type 2 processes (plus the possibility of some rare exceptions).

The crossover problem is much more serious, I believe, than Samuels claims. Not only are there processes exhibiting characteristics from each cluster of properties, but there are processes exhibiting S2 / Type 2 characteristics at first but that, with practice, repetition, etc., *acquire* S1 / Type 1 characteristics. While the 'Types' account allows us to characterize such changes, it does not explain how the transition is possible, or how one process can *change* in any way (or, worse, how a process can change according to one characteristic, but not the other). In a similar fashion, Stanovich (2004, 38 & 42) acknowledges the possibility that a process may change type, but does not provide a framework in which we can specify and understand what exactly is changing, and what is going on at the phenomenological, the

---

[42] The most known example of an experiment demonstrating the Stroop effect is when the experimenter asks the subjects to name the color in which a word is written and that the word presented is the name of another color (e.g. the word 'blue' written in green). These experiments suggest that reading abilities are as automatic as perceiving the color, because the word (automatically) read by the subjects interferes importantly with their reaction time when they have to name the colors they see. This 'interference' is what is meant by 'Stroop effect.'

representational or the neurological level. I will explore how to think about what Karmiloff-Smith (1992) calls 'modularization' in 3.3.2.

The third problem Samuels discusses is the unity problem. What are the unifying characteristics of Type 1 and Type 2 processes (if any)? Sloman's (1996) suggestion is to distinguish between rule-based and association-based processes, which would correspond roughly to the distinction between classic computationalism and connectionism; Carruthers' (2006; 2009) proposal is to understand S2 / Type 2 processes as being realized in cycles of S1 / Type 1 processes (as I examined, and rejected, in 2.2) and, finally, Evans (2008; 2009) argues the distinction between Type 1 and Type 2 is linked with how working memory functions.

While this suggestion by Evans might be the most interesting one (Samuels, 2009, 144), I think it would be very surprising for the Type 1 / Type 2 distinction to be *one*-dimensional, even if some of the dimensions might be more characteristic of one type than the other (cf. 3.2.1). In the following sections, I will detail my own account of dual-process theories: I think that, while the 'Types' account is interesting, it needs some rethinking.

3.2 Continua, conceptual space, frontiers and some grey areas

What we saw thus far might give the impression, and has given many the impression (e.g. Machery, 2009), that it is futile to try to sort cognitive processes in one of two (or three, four) precisely defined classes. Dual-process theories look like a wild goose chase: it seems as if there is no clear difference between S1 and S2, or even a clear distinction to make between two types of processes. Machery's (2009, chapter 5; forthcoming) critique of Stanovich (1999; Stanovich and West, 2000) could be made to many dual-process theories, even more recent ones:

> Like many dual-process theories, Stanovich's dual-process theory is somewhat unsatisfying. The cognitive processes that are assumed to constitute System 1 and System 2 are not described in any detail. Their triggering conditions and the nature of the integrative or non-integrative mechanisms are left pretty much unspecified. As a result, it is difficult to derive any clear predictions from his theory, which is better suited to provide post-hoc explanations. Thus, in spite of the real interest of Stanovich's work, his dual-process theory illustrates the pitfalls to be avoided in building a multi-process theory. (Machery, 2009, 147-148)

I believe this impression is correct (see also Keren and Schul, 2009), but only to a certain extent. Behind many accounts of dual-process theories, there is the assumption that we will find two neatly distinguishable systems, groups of processes or kinds of processes *able to explain all of cognition*. This is true of all of the authors discussed so far in this dissertation. The 'Systems' view is well illustrated by Lieberman (2007; 2009), and Stanovich (2009) and Evans (2009) are adopting a 'Types' view. These views are, according to what I have been arguing, unable to account for some phenomena, and they should be abandoned in favor of a more comprehensive account[43].

However, Carruthers' (2006; 2009) account – eliminating any deep functional difference between the systems or processes – is not a very plausible alternative. While his account is interesting, he too has problematic assumptions, such as that the characteristics attributed

[43] This does not mean in any way that their data is useless and that they did not make any useful empirical hypotheses. I believe their work only has to be reexamined under a new light and that the perspective developed in the following sections might help to understand some of the difficulties they encounter.

to S2 would be uniquely human[44], as well as the very fact that he does not recognize there are deep functional differences between cognitive processes (cf. 2.2). Also, his account is not precise enough – as we saw in 1.3, his definition of the notion of 'module' is open to criticism – and his theoretical framework has limited power to account for some of the cases studied in the literature. Again: the solution is not to abandon completely the current dual-process frameworks, as even Carruthers' brings forward an important and interesting idea – the role of iterated Type 1 processes and a plausible account of (some of) consciousness – by reinterpreting Baars' (1988; 1997) general workspace theory in the dual-process framework.

All of the theories mentioned above only need to be placed in a wider framework where we can understand that they only explain parts of cognition, and this has been the driving force behind most of the arguments presented in this dissertation against diverse dual-process accounts. Samuels' (2009) proposal was seen as the most plausible up to now, but it also has

---

[44] In a framework where, as I will argue, we can consider independently each characteristic attributed to S2 / Type 2 processes, there is no reason to think Type 2 processes are uniquely human, or that they are necessarily linked to language. Of course, language plays a very important role in human cognition (and it certainly enhances a lot of our abilities), but it does not mean that all of the abilities we have that are linked to language *need* language to exist. For example, language can help and enhance our capacity to learn, but it does not mean that, without language, there would be no learning.

Taking S2 / Type 2 processes to be uniquely human is an assumption made by many models and it certainly is wrong. As Evans explains: "Taken in conjunction with the evidence of higher-order control systems in animals (Toates 2006), these arguments [about the possible existence of explicit memory systems in animals] suggest that dual-system theorists would be better off claiming that System 2 cognition is uniquely developed, rather than uniquely present, in modern humans. Such an argument also has much greater evolutionary plausibility." (Evans, 2008, 260)

This distinction between 'uniquely human' and 'uniquely developed in humans' might have very important consequences on how we link dual-process theorizing with reflections on the evolution of language. Many Type 2 processes are indeed strongly linked to capacities necessary for developing language, and being able to distinguish each of these capacities might prove to be crucial: we can think of Type 2 processes that are nonlinguistic. As Fitch mentions, after his discussion of animal cognition: "[...] a large body of experimental work demonstrates considerable cognitive abilities in nonhuman animals. Many different vertebrates have a surprisingly rich conceptual world and a broadly shared cognitive "toolkit" (Hauser, 2000), and the data reviewed above leave little doubt that sophisticated cognition is possible in the absence of language. Many capabilities that were long thought to be unique to humans have now been demonstrated convincingly in animals. These include cross-modal association, episodic memory, anticipatory cognition, gaze following, basic theory of mind, tool use, and tool construction." (Fitch, 2010, 171-172)

important problems, as the types of processes, if defined with the same characteristics (even if we use just a few of them) as those attributed to each system in Table 2.2, will exclude many actual cognitive processes (e.g. those Evans calls Type 3). Samuels claims this problem is not serious since he doubts that a few exceptions will prove to be fatal for the 'Types' account of dual-process theories[45]. His view would be plausible if there were *only a few* exceptions.

As it turns out, most processes actually *are* exceptions in the sense that most cognitive processes do not neatly comply with the lists presented in Tables 2.2 and 2.3 for at least one characteristic. 'Searching' for an answer in our memory surely has a controlled and conscious element to it, but most of it is resolved automatically and unconsciously[46] (to use Baars' expression: "That Little Pause before the Answer Comes to Mind"; Baars, 1997, 49). Reasoning is also mostly controlled, although the heuristics discussed previously play an important role. Even seemingly completely automatic processes, reflexes for example, usually have a few 'Type 2' characteristics or, at least, can acquire them. As Machery recently puts it:

> I suspect that the slow / fast, conscious / unconscious, linguistic / nonlinguistic, recent / ancient, non-heuristic / heuristic, rule-based / similarity-driven, etc. dichotomies are by no means aligned. That is, I argue that these dichotomies are orthogonal from one another: some processes are slow and nonlinguistic, some processes are heuristic, linguistic and conscious, some processes are ancient and non-heuristic, and so on. [...] Some non-automatic processes are ancient and nonlinguistic (Evans, 2008), some automatic processes are not heuristic (e.g. Woodward and Allman, 2008), some automatic processes (such as the abilities of experts) are acquired by learning [e.g. Dreyfus and Dreyfus, 1986], and so on. (Machery, forthcoming, my translation)

---

[45] "[...] should the existence of crossovers lead us to reject dual-process theory? The answer is, I maintain, that it poses no serious problem, so long as crossovers *are not too numerous or too extreme*." (Samuels, 2009, 140, my emphasis)

[46] This is a very good example of multiple iterations of S1 processes that can become conscious, as Carruthers suggests.

Could it only depend on how we individuate the processes? This might be, of course, part of the answer, but, in general, these 'Types' have more the feel of an archetypal description of 'ideal' processes than one of an accurate description of actual processes. It seems the characteristics were attributed to each process of a given type *a priori* with an intent to oppose characteristics to obtain a perfect list of opposing traits and a couple of examples in mind – input processes most notably (Fodor, 1983) –, but they fail to encompass the diversity of the cognitive processes – everything that would fall in between 'input processes' and 'higher cognition' in Fodor's account. This impression comes from the many problems identified thus far.

3.2.1 One continuum or multiple continua?

Following an insight along the lines of Hammond's (1996) suggestion, interpreting the distinction we are looking for as a continuum between 'pure Type 1' and 'pure Type 2' processes has the potential of solving some of the difficulties encountered. As Osman explains:

> Hammond's (1996) cognitive continuum theory proposes that different forms of cognition (intuitive, analytical, common sense) are situated in relation to one another along a continuum that places intuitive processing at one end and analytical processing at the other. (Osman, 2004, 992-993)

Yet, such a proposal also raises new questions and, eventually, new difficulties.

A first problem that comes to mind is that the characteristics listed in Table 2.2 are taken to be causally related (e.g. it is assumed each type forms a natural kind, as defined by Boyd, 1990; 1991; cf. 3.2.4). This is opposed to the very idea of a continuum, as it distorts the central idea of dual-process theories. Yet, as we saw earlier, dual-process theories as they currently stand cannot explain all cognitive processes: while they are very useful to understand and explain a part of the mind, a large part of it cannot be accounted for. As mentioned, the crossover problem – that some processes have characteristics from both columns of Tables 2.2 and 2.3 – is unavoidable.

A clear advantage of the idea of a continuum is that it allows this mix of some of the characteristics of Tables 2.2 and 2.3 in order to characterize correctly a controlled mechanism demanding low effort, or a domain general unconscious process, if such mechanisms were to be identified. We would not need to classify it in either rigid version of Type 1 or 2 – thus eliminating the need of a new category such as 'Type 3 processes' to characterize the control of attention as in Evans' (2009) account, or more generally a category defined with characteristics from both columns.

Dual-process theories, in an account based on a continuum, would help us identify these archetypes correctly and allow us to classify cognitive processes as being far or close to each of these archetypes, with a precise description of each process. We would then be able to classify correctly processes from one extreme to another – from a perfectly autonomous process like perceiving shapes to a rather reflective process like understanding a complex philosophical paper. Researchers could then have a tool allowing for a better understanding of crossovers, without having to assume that exceptions are rare (as does Samuels, 2009, 140), as well as processes that are badly understood when forced into either side of the dual-process distinction.

This perspective does not undermine the relevance of dual-process theories: it might even help us to reconcile many approaches to dual-process theories that might diverge only in appearance, e.g. because they put their emphasis on one element rather than another (e.g. Stanovich, 2009 and Evans, 2009 suggesting a different 'third' type or system). The idea I advance here is that processes truly dual (processes that, when compared, have opposed characteristics) would still be opposed (each at one extreme of the continuum), and the 'Continuum' framework would thus preserve the central intuitions of current dual-process theories. Placed in such a 'Continuum' account, it would also be possible to characterize, without losing the explanatory power of the very useful frameworks provided by dual-process theories, the processes that are hard to identify with one of the two lists of characteristics. I believe that the theories suggesting modifications to the initial forms of dual-process theories, such as Stanovich's (both 2004 and 2009) or Evans' (2009) might end up as two strongly similar theories in this 'Continuum' account. The existence of 'in-between' processes also highlights some of the difficulties encountered by

Lieberman's (2007; 2009) account, viz. that even if many characteristics are strongly causally connected, there is evidence that it is not true for all of the characteristics presented in Table 2.3 (e.g. Stanovich's, 2009 distinction between the autonomous and the reflective mind is to explain why some processes he attributed – in 2004 – to S2 do not correlate with general intelligence, while it is a characteristic assumed to be linked with S2 in many accounts, cf. Table 2.2).

However, this idea of a continuum has an important problem: classifying processes along a continuum may be a nice idea in theory, but in practice we have to explain how we should proceed to this classification, and justify our choice. It is not as easy as it may seem. Since most processes are nonstandard with respect to the Type 1 / Type 2 distinction, and vary on different characteristics (e.g. the difference is not only about consciousness, viz., they do not necessarily vary, at once, on *all* characteristics attributed to a given 'Type'), determining which properties has what effect on the classification is far from being straightforward. Should we posit that two processes that diverge on only one characteristic at the same position in the continuum, or should we consider that there is one characteristic (e.g. automaticity) more influential than another (e.g. domain specificity) on whether one process is closer to archetypal Type 1 processes? This problem is, I believe, impossible to solve in a framework where we only include a one dimensional continuum and that is why we have, ultimately, to reject Hammond's (1996) proposal.

However, this problem can be easily solved by posing a conceptual space, a set of perpendicular continua. *Each* characteristic has different effects and an interesting and heuristic representation of this complexity is conceptual spaces (Gärdenfors, 2000[47]). My proposal is that the characteristics presented in Table 2.2 should all be understood as continua and that, together, they form an n-dimensional conceptual space, where n is the number of characteristics considered (the number of continua forming the conceptual space).

---

[47] In his book, Gärdenfors uses conceptual spaces to model representations. My goal is different, but Gärdenfors' proposal, to use "conceptual spaces as a framework for representations" (2000, 4), is interesting to conceive how to describe, represent and classify cognitive processes.

In such a conceptual space, we can ascribe a position to the cognitive processes that are studied. Once all processes are posited in the conceptual space, we might find that they form identifiable clusters, reflecting but not limited to the various types that have been proposed. I will briefly discuss the value we should attribute to *n*, the importance of frontiers and clusters in this conceptual space and the place of the Type 1 / Type 2 distinction in my framework. I will then consider some advantages of my view before examining what could arguably be considered the most controversial characteristics of Table 2.2.

3.2.2 What is *n*'s value?

The first thing to consider before we elaborate this kind of conceptual space is the number of dimensions that researchers will need to take into account. The view I present here is tentative as I do not have a clear answer to this question. The answer will, mostly, come from empirical studies – as it will depend of whether we can, or not, distinguish a given characteristic from the other: only being able to do this distinction conceptually is not enough.

My view does not reject the possibility that some characteristics co-occur most of the time, and it does not undermine the importance and relevance of some co-varying characteristics: the idea that characteristics must be distinguished from one another rather means that they should be distinguishable at least in principle, but preferably[48] in actual, verifiable, situations. Is the unconscious character of a process the same thing as its implicit functioning? Is being the default process the same as being automatic, and are those characteristics necessary in order to be fast? Are all processes that are independent of general intelligence also independent of working memory? We can distinguish most of these concepts from one another, but it may be possible that, in practice, they are the same characteristic that, because of current ignorance, we identify with two different labels. There

[48] It is conceivable, but not desirable, that in a given case we might have very good reasons to preserve a distinction between two characteristics co-varying perfectly.

are dozens of questions like these we will need to answer before having an appropriate and useful description of the processes, some of them conceptual and a priori, but most empirical.

However, it is important to note that specifying what are the 'correct' characteristics we should consider is, in this framework, less crucial than it was in Samuels' account. He submits the problem this way:

> Should we invoke, for example, a distinction between parallel and serial processes, as Sloman does (1996), or should we resist this suggestion, as Evans does (2008)? *Mutatis mutandis* for conscious versus unconscious, controlled versus automatic, evolutionarily ancient versus novel, and so on. (Samuels, 2009, 139)

For Samuels, this identification of the correct characteristics was an important problem with potentially significant consequences. Excluding or including one set of characteristics instead of the other could radically change the proposed account of these 'Types' – for instance, considering parallelism as a criteria for S1 or Type 1 potentially excluded some processes from the system or the type. This is not the case in a conceptual space account, as a redundant dimension would not change much in the clustering of processes and, if it did have effects on the clusters, it is either a sign we undervalued the importance a characteristic has for a certain set of processes, or that the characteristic should be excluded because it is inappropriate (e.g. it would be the case with evolutionary recency according to Samuels, 2009, 140).

Thus, if we examine redundant dimensions in my framework, we should be able to identify them (in the sense of making them one) as every process examined will plot at 'the same place' on the two respective axes; we would get a perfect correlation[49,50]. It might also turn out that some characteristics are less interesting – for example, they might have less

---

[49] This is the main idea behind principal component analysis: highly correlated dimensions (components) are fused to leave only the set of maximally uncorrelated (i.e. orthogonal) dimensions.

[50] There is a simple algorithm we could begin research with: starting with Table 2.2, as elaborated by Evans (2008), or something similar, we determine which dimensions correlate strongly and merge them and, when new characteristics appear in the literature, we add new dimensions to the space. The iteration of this algorithm offers a construct and manages the space. Thanks to Pierre Poirier for this suggestion.

explanatory power – than others that better capture some aspects of the processes investigated, and these could be eliminated as well. I will elaborate on this point in 3.2.4 and in 3.3.3.

I believe identifying the set of the most important characteristics will be made easier with this account of dual-process theories, more so if there are systematic examinations of the characteristics. Maybe some axis will be binary, but most will, I believe, be true continua. For instance, automaticity is not a question of being automatic or not, while some other processes are partly automatic, some others are automatic, but just sometimes (e.g. the control of breathing). Since there is at least three possibilities for the 'automatic / controlled' characterization, a continuum will offer a more accurate description of the analyzed process. It is important to note that, in these cases, plotting a process at a certain point of a given continuum is not an easy task, but with more and more processes placed in the conceptual space, the comparisons between diverse processes will become easier to do. The idea is not that the place of a process on a continuum is an absolute characterization (e.g. 'Process X is 77% conscious') that identifies properties, as it would be impossible to determine. The idea advanced here is rather that it is a heuristic classification enabling us to understand better the process in question in relation with other cognitive processes ('This process should be at 77% on the consciousness axis as it is more than x but less than y'). A similar move is made by 'traditional' dual-process theories: input modules, in Fodor (1983), had certain characteristics only useful when comparing them to higher cognition (e.g. 'being fast' is a characteristic we can only attribute if there are some slow processes), and similar characterizations are found in other accounts of the two systems or types of processes.

Given a number of dimensions – that I will leave undetermined in this dissertation – in which we place and compare different processes, the important question now becomes: does the distinction between Type 1 and Type 2 processes still has a purpose?

3.2.3 Frontiers and clusters

Multidimensional spaces are hard to grasp and their complexity is impossible to represent (completely and clearly) on paper. However, there are a number of mathematical tools that can be used to represent them and use them in a useful and intelligible way. I do not wish to enter in the details such an account could have: my goal is only to sketch the general idea of what a conceptual space of processes could look like, and how it could be useful to further current research on dual-process theories. Most of the ideas related to prototypes here are inspired by the connectionist framework (Bechtel and Abrahamsen, 2001; Eliasmith and Anderson, 2002) as there are very useful tools in this framework. Figure 3.1, adapted from an illustration of a learned partition of a hidden-unit activation-vector space (Churchland, 1989, 169 & 203), is an example of the kind of conceptual space that could result of the consideration of three characteristics (e.g. speed, automaticity and parallelism).

Each dimension is the continuum of a given characteristic, each prototype process is the center of a cluster of processes and each side of the divider in the center would mean the process is classified as either (Proto)Type 1 or as (Proto)Type 2. Of course, the actual conceptual space that would result from the consideration of automaticity, speed and conscious character is most probably very different from Figure 3.1, but this is used only to illustrate my idea.

There is a strong chance, according to the current evidence we have (Evans, 2008; Evans and Frankish, 2009; Stanovich, 2009; Toates, 2006) that clusters would form in this space, viz. most processes would be grouped together when they have many characteristics of Type 1 or many characteristics of Type 2 in common. However, exceptions would still be possible and they would occupy other positions in the conceptual space: but we would still be able to classify them as Type 1 or 2, if necessary. We might even find that there are two (or more) types of Type 1 processes (e.g. the old and evolutionary, and the recent and modularized, cf. 3.3.2). The main advantage of this view is that it preserves the specificities of each process, but that we can still qualify it as being of a given type if necessary (cf. 3.3.1). And it lets determination of the amount and organization of 'Type-structure' to be determined empirically.

Even if a process is not precisely in the area occupied by the prototype, the conceptual space allows us to simplify for the means of explanation, or to sketch a quick portrait of the situation when needed. The divider at the center of Figure 3.1 represents the frontier between Type 1 and Type 2: on one side it would be classified as Type 1 by the algorithm, and it would be described as Type 2 in the second half (cf. 3.2.4 for a more thorough description). The divider is not a plane: the distinction between Type 1 and Type 2 will probably not be linear, as it will depend of each of the characteristics considered. We should be able, in such a conceptual space, to determine if one characteristic has more impact than another on whether a process should be considered to be Type 1 or Type 2.

This frontier however is not absolute. A process that appears clearly on the divider or near it would be partly Type 1, partly Type 2 in almost equal respects. Should we then force the process in either one of the categories? Why would we? What is clear is that the distinction between the types of processes is not binary. We can easily imagine the difference between types as a sigmoid function (cf. Figure 3.2) in which the magnitude of the curve will only be revealed by more research. We can also suppose there will be a grey area near the divider, an area in which a process would be categorized as both, or neither type. Moreover, the division between types in this grey zone, as many distinctions made in fuzzy systems, might also depend on the relations of the analyzed process with other processes and with the environment. For example, when categorizing a Type 2 process that is becoming more and more automatized, it might be preferable to compare it to other acquired Type 1 processes.

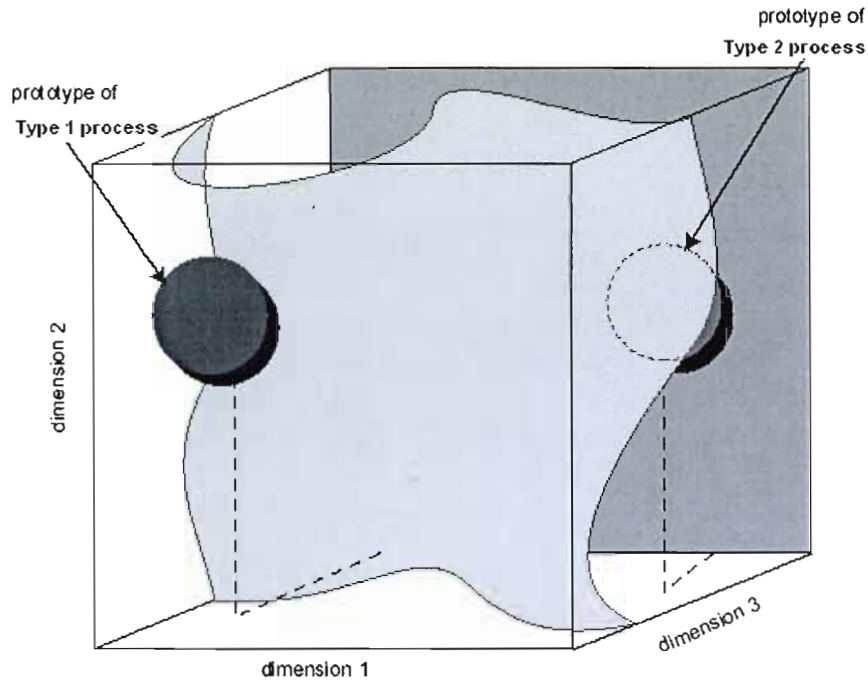

Figure 3.1 Prototypes of Type 1 and 2 processes and frontier in a three-dimensional conceptual space. (Adapted from Bickle, Mandik and Landreth, 2010, Figure 3.)

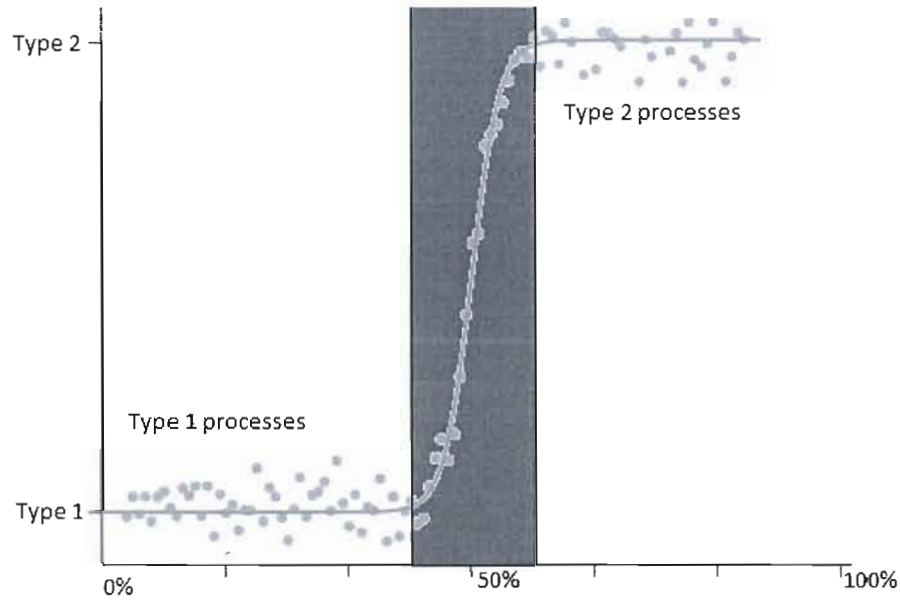

Figure 3.2 Illustration of a projection of the position of various processes (one shown) along a given continuum (e.g. automatic / controlled, where 0% is not controlled at all, and 100% is fully controlled) in the conceptual space with their classification as Type 1 or Type 2 process, showing the grey zone between each type.

3.2.4 Type 1 and Type 2 as a heuristic distinction

If the framework presented above is accepted, the Type 1 / Type 2 distinction acquires a new meaning (or, less controversially, it is at the very least represented in a different way). From a characterization of processes, it becomes the localization of a process in a complex conceptual space, divided in (at least) two regions. These regions are more or less defined: depending on the characteristics we find once the conceptual space is elaborated and we describe the clusters forming in this conceptual space. It is not clear however *what* a given region of the conceptual space identifies[51].

The problem we face here is one of identification of natural kinds. Does the Type 1 and Type 2 processes distinction (whether it is in my conceptual space account or in Samuels', 2009 account of dual-process theories) map, or not, to a distinction we can reliably use in psychology and in neuroscience?

To use Machery's characterization of natural kinds – inspired from Boyd's (1990; 1991) –, natural kinds are objects sharing a large set of properties, they possess these properties because of some causal mechanisms and this set of properties has to be specific to these objects (adapted from Machery, 2009, 241). Thus, the two[52] questions we must answer affirmatively to identify a natural kind are the following: "Are there clusters forming?" and "Do these clusters have many characteristics in common because of a causal mechanism?" In the previous sections, I argued that, while we can probably answer the first question positively, although there might be more than just two clusters, it remains to be seen if this is also the case for the second question.

---

[51] It will also vary if there are more than two regions identified, for instance, if we were to find four very useful clusters.

[52] The specificity criterion is less interesting in this context. Still, it will be satisfied automatically by the conceptual space itself, as we will be able to see it if there are two processes sharing the exact same set of properties.

We can offer no affirmative answer for the 'Systems' accounts, nor should we, at least for the moment, expect an affirmative answer for Type 1 and Type 2 processes[53]. If many processes form a cluster, we will be able to search for the underlying causal mechanism of these processes, and the chances are the underlying causal mechanism will vary from one cluster to the other – for instance, Fodor-like input modules bringing about a representation and a reasoning process producing a particular intuition might involve very different causal mechanisms. In any case, we must understand each process and what role it plays in our cognitive architecture.

While elaborating the conceptual space I suggest, we might find out that the distinction between Type 1 and Type 2 processes identifies a natural kind, or not. In the first case, my framework is useful to see the relations between each type and to characterize properly each process according to its specificities and mapping them in the appropriate region of the conceptual space (but, as I argued in 3.1; I think it is unlikely that an account similar to Samuels', 2009 has the potential to explain all cognitive processes, mostly because of the problem of crossover characteristics, cf. 3.1 and 3.3.1). It can, as well, be useful to refine the characteristics and the criteria used. I suspect however that it will not be the case that the Type 1 / Type 2 distinction 'carves the nature at its joints.'

However, if the distinction does not identify natural kinds, should we reject it completely? As I emphasized through 2.2, I think Carruthers' (2006; 2009) account does not go in the right direction because of the problems I identified and objections I raised to his too weak notion of module and to his account of the realization of S2. Still, Carruthers identifies some important properties of at least some Type 2 processes: there are certainly some of these processes realized in multiple iterations of Type 1 processes, and action-rehearsal most probably has a role to play in some of them. It is not, however, the whole story.

---

[53] The case for or against this Type 1 / Type 2 distinction being a natural kind is certainly not complete, but I did cast some doubts on it. If Type 1 and Type 2 processes identify natural kinds, it will become clear with the conceptual space. This framework will help us find the shared properties *and* the causal mechanism(s).

An elimination of the distinction between Type 1 and Type 2 processes could be desirable if we were to find no clustering at all in the conceptual space. However, as I mentioned previously, it is almost certain that we will find *some* clustering as many processes are alike. Marcus (2004) offers even more evidence with his suggestion that similar neural structures are often found as genes replicate, at least initially in the course of evolution, the same structures over and over. Another reason in favor of keeping the distinction is important to invoke: the models offered in some accounts of dual-process theories such as Evans' or Stanovich's models in psychology of reasoning, or Lieberman's very rich and interesting framework (cf. Chapter II), "has generated a wide array of fruitful explanations for many phenomena." (Samuels, 2009, 138)

The difficulties encountered by these models are somewhat small compared to their advantages, and the goal of my suggestion is to provide tools to amend current dual-process theories and ease the resolution of some of the difficulties identified, and finally to get to the bottom of some of the problems of these theories (and, hopefully, find a way to rectify these frameworks). I identified many such problems and difficulties in this dissertation, and I am certain there are many more to discover, as some accounts (e.g. Lieberman's, Samuels') are still in their infancy.

I must emphasize the point I made in the last paragraph, as it is a potential source of misunderstanding. I do not argue for the elimination of notions such as 'module,' or 'Type 1' and 'Type 2' processes, although my framework provides the tools to add precision to the definition of these notions, or even to eliminate it, if needed. The model suggested here is just a different, and I believe better, way to conceptualize how and why some processes are very similar, and how and why they differ from other cognitive processes. Discussions about modules are very fruitful, and Fodor's (1983) work opened the door to many inquiries. These discussions, and many others related to dual-process theories, also gave us very important tools to analyze some empirical results: it would be a tremendous error to throw out the proverbial baby with the bath water.

In my view, if we integrate the conceptual space to our thinking about dual-process theories, it will, at first, bring us to consider the distinction between 'Type 1' and 'Type 2'

processes as *only* a heuristic distinction, viz. a distinction that has no other purposes than being useful in designing experiments and interpreting the results. In other words, not only does this distinction is not taken to identify natural kinds, but this distinction gives rise (or, at least, has the potential to cause) major confusions and errors. One of the most striking one, I believe, is the potential conflation of Fodor-modules, or processes that are very close to what is identified with this notion, with skills acquired by experts and with perceptual-motor tasks, that – after a lot of training – become very much alike modules (cf. Dreyfus and Dreyfus, 1986; Fitts and Posner, 1967; Karmiloff-Smith, 1992). I will return on this important distinction in 3.3.2.

The key point here is that I do not claim there is only one type of 'Type 1' processes, nor do I claim there is only one type of 'Type 2' processes. Maybe some important notions will be confirmed – I believe Fodor-modules is one very likely example of processes that will cluster strongly in the conceptual space –, maybe some others will be replaced, etc. It might also give evidence for (or against) Carruthers' idea of S1 and S2 having no deep functional differences. We might in effect see some of the S2 processes cluster in a region that *can* be reduced to the iterated activation of S1 processes. Maybe the conceptual space will confirm that some processes appearing to be exceptions not possible to include in some of the frameworks presented in this dissertation are not really exceptions after all. As I mentioned at the beginning of 3.2, only more research (and using the conceptual space!) can confirm or infirm much of the claims made by dual-process theories.

This framework also has other important advantages that I will describe in the next section.

3.3 Some advantages of this view

I see at least six major advantages to my view over the current Type 1 / Type 2 dichotomy (as in Samuels, 2009). First, the categories we use would be justified by the position in the conceptual space of the processes we regroup under them. The fact that the

processes have more or less the same position in the conceptual space and that their position differs from the emplacement of other processes in the conceptual space (cf. 3.2.4) illustrates how similar these processes are between them and how they differ from other processes, not comprised under the same category. Second, Type 1 / Type 2 will not be seen as a dichotomy anymore. The *dual* character of some processes explained by dual-process theories would remain, as there would most probably be similar dualities than those already identified in the literature (as in Table 2.2), but some intermediate processes can be allowed (cf. 3.3.1). The 'Continuum' account and the conceptual space do not undermine the explanatory power of dual-process theories, but these two accounts place the current accounts in a wider framework, a framework where it is possible to explain the processes that are not explained by dual-process theories. The distinction between Type 1 and Type 2 processes in the suggested conceptual space is not 'either-or' anymore, a requirement of dual-process theories that prevents the explanation of many processes. A third advantage, related to the second one, is that the 'two systems' used in the current literature are, most probably, no longer seen as the only two notions to explain the entire mind (all of its processes). It puts an end to the binary and oversimplified dual-process theories (cf. Evans, 2008, 270), some of which have a central role in the literature (cf. Chapter II). Fourth, as mentioned at the end of last section, my framework allows an explanation of acquisition and automatization of some skills after much training and repetition (cf. 3.3.2). Some processes are initially 'Type 2' but they gradually acquire characteristics attributed to Type 1 processes, such as automaticity and the possibility that they function unconsciously (e.g. riding a bicycle, playing chess, reading, understanding a language, etc.). It is also possible for a processus to go in the other direction, viz. from Type 2 to Type 1 (*de*modularization). Fifth, the conceptual space will allow us to verify and elaborate new and interesting hypotheses – by manipulating the conceptual space, for example by adding or removing characteristics (i.e. dimensions) to see the effect, researchers will be able to have a better grasp of how the mind works (cf. 3.3.3). Lastly, while the structure offered by dual-process theories, whether a given theory uses the 'S1 / S2' or the 'Type 1 / Type 2' distinction, brings researcher to make unjustified assumptions such as believing that a controlled process will be linked to language and general intelligence while it is sometimes not the case (Stanovich, 2009; Stanovich and West, 2008), the

conceptual space account would allow researchers to see in details which characteristics are co-varying and which do not.

3.3.1 Crossovers: mixing characteristics

With the characteristics listed in Tables 2.2 and 2.3 taken individually, and not seen as a set of *necessarily* co-varying properties (even if many of those properties often co-vary), it is possible, with the conceptual space, to understand better some processes that have characteristics from both columns (a process encapsulated but slow, etc.). Moreover, it is possible to preserve (and maybe even refine) the explanations of the traditional dual-process theories or those provided by frameworks such as the Evolutionary Psychologists'.

This point is crucial, and it is hard to understand why Samuels (2009) argues that 'crossover' processes, viz. processes having characteristics that are in the first column and characteristics that are in the second one at the same time, are rare. Whether they are rare or not is not the point that should be made: such processes exist and even though their number might be small (maybe it is not), they are an important part of cognition. For example, Evans' proposed Type 3 is essential for the coherence of many of the dual-process theories themselves (cf. Evans, 2007; Evans, 2009, 47-48). Furthermore, as we have seen with Machery's quote earlier (p. 64), the existence of processes that have crossover characteristics is one of the reasons why many researchers in cognitive science reject the extremely useful and rich framework offered by the many dual-process theories (or why they only adopt a view using *some* of the intuitions behind dual-process theories, e.g. when de Sousa, 2010 refers to this literature with the expression 'two-track mind').

However, some might see my proposition as a departure from dual-process theories because, as we have seen (cf. 2.1), current dual-process theories share two tenets: that the distinctions made in Tables 2.2 and 2.3 align and that a given process is either part of S1 or of S2 (or, in Samuels' account, either Type 1 or Type 2). My view goes against this 'canon' on both counts. I do not see this as a problem since, as I argued thus far, accounts sharing

these principles are unable to account for some cognitive processes. But why still call it a dual-process theory?

Samuels' appeal to *explanatory keep* holds part of the answer. My account does not restrain the current theories in any way: it only places them in a larger and richer framework. It is thus possible to retain the advantages and explanations offered by Stanovich's or Evans' accounts, but evade much of the problems they encounter (at least, those I identified here). In this way, it is possible to keep the explanation of the processes *that are* dual, like some phenomena discussed in the psychology of reasoning, and have room to explain processes that do not fit in a framework inferred from data coming from these experimental paradigms (e.g. Kahneman and Frederick, 2002; 2005).

3.3.2 The modularization of Type 2 processes

Expertise is a wide topic I cannot fully cover in these pages. However, I need to say how it could be accounted for in the conceptual space that I am suggesting here – especially because *it can be*. Many dual-process theories mention expertise and the automatization of some processes, but these authors offer no explanation (or, at least, no detailed account) in their framework for how the acquisition of new characteristics or how the shift from one side of Tables 2.2 and 2.3 to the other happens. For example, Stanovich states that "processes can *acquire* the property of autonomy" (Stanovich, 2004, 38) and he illustrates this phenomenon a few pages later with the Stroop effect. Yet, he offers no explanation of how this fits in his dual-process account.

What I claim here is a bit stronger: if dual-process theorists wanted to give a fully detailed account of the automatization of some processes they would encounter insurmountable difficulties that they could only resolve by modifying some of their basic assumptions. The first difficulty these authors would come across is that it is not possible in their framework to explain why there is a change in one characteristic but not necessarily in the others. Remember that in current dual-process theories, characteristics are assumed to align. In fact, not all characteristics change at the same time or at the same rate, and some

never change when one is learning and repeating a new ability. The second difficulty would be to account for the gradual nature of these changes, since a process cannot be partly Type 1 and partly Type 2: it can only be either one. Some of the processes 'in transition' have in this sense a curious behavior current dual-process theories cannot explain (cf. Dreyfus and Dreyfus, 1986, chapter 1 for the first two difficulties). A third difficulty is mentioned by Samuels (2009, 140) when he discusses which, if any, characteristic should be dropped from Table 2.2: as he explains, the characteristic of evolutionary recency is incompatible with the idea of such an automatized Type 2 process like reading (while it is a much more useful notion in the context of thinking about Type 1 processes as using heuristics). Reading is, at first, very hard and complex to execute. A child who just learned how to read has to make much effort to *understand* what s/he is reading (or imagine you were learning to read Chinese characters). Decoding each syllable (or ideogram) takes all of the attention, as it is executed in a serial manner, and there is no attention left to decode what meaning lies behind these syllables. With time, decoding the syllables becomes easier and a proficient reader will execute this task just as easily as s/he perceives the written word and its color (without 'decoding' it), thus explaining much of the Stroop effect. But this acquisition of reading skills does not fit neatly in current dual-process theories[54].

Karmiloff-Smith (1992), in her book revisiting the notion of 'Fodor-modules' presents what she calls 'modularization,' and her description is very close to what I have in mind here. She gives the example of someone learning to play the piano:

---

[54] And it remains true even if we drop evolutionary recency as the first two problems remain. However, most current dual-process theories use evolution as an important factor in their classification (e.g. Carruthers, 2006; Stanovich 2004).

> When one is learning to play the piano, initially there is a period during which a sequence of separate notes is laboriously practiced. This is followed by a period during which chunks of several notes are played together as blocks, until finally the whole piece can be played more or less automatically. It is something like this that I shall subsequently call "reaching behavioral mastery." But the automaticity is constrained by the fact that the learner can neither start in the middle of the piece nor play variations on a theme [...] It is only after a period of behavioral mastery that the pianist can generate variations on a theme, [...] The end result is representational flexibility and control, which allows for creativity. Also important is the fact that the earlier proceduralized capacity is not lost: for certain goals, the pianist can call on the automatic skill; for others, he or she calls on the more explicit representations that allow for flexibility and creativity. (Karmiloff-Smith, 1992, 16)

This description parallels the acquisition of many other skills, including writing and reading for example, but also chess playing, the acquisition of expertise in a workplace (it accounts for the acquisition of skills from that of chicken-sexers, cf. Pritchard, 2005, to the skills of physicians and nurses), driving a car or a bicycle, and so on[55]. Karmiloff-Smith here identifies only a few stages, but Dreyfus and Dreyfus (1986) identified five levels of proficiency: novice, advanced beginner, competent, proficient and expert (Fitts and Posner, 1967 presented three such stages). These stages, used to explain intellectual skills or functions also parallel very well how perceptual-motor expertise is acquired (Rosenbaum, Augustyn, Cohen and Jax, 2006), viz. actions we make everyday such as picking objects up, but also other actions that seem very easy to us like walking, moving the lips and tongue to

---

[55] For Karmiloff-Smith (1992) all modules are 'constructed' in this fashion; according to her account, only attentional biases can explain how modules are developed. This is an interesting claim, but it is not one I have to endorse or to reject: my suggestion is compatible with both accounts.

If there are innate modules (e.g. 'true' Type I processes), as it is usually understood by Evolutionary Psychologists, these would most probably appear in the conceptual space as a different cluster. If Karmiloff-Smith is right however, my prediction is that there would be no differences between modules assumed to be innate and those assumed to be acquired.

produce a given sound, etc., suggesting that a similar transition from Type 2 to Type 1[56] occurs across many cognitive processes.

While this description seems very hard to account for in current dual-process theories[57], it fits well in my framework: it allows that some of the characteristics of a process can change in given circumstances. As the process modularizes, we only have to examine which characteristic(s) changes and map the process anew. We can then follow the evolution of the skill's mastery. These processes, initially Type 2 (as defined in the conceptual space), gradually and continuously acquire more and more Type 1 characteristics. I believe the conceptual space can be a very interesting and rich way to interpret the analysis made by Dreyfus and Dreyfus (1986), but this is ground for further research. What is important here is that the framework I suggest offers ground for reinterpreting the research already done, and opens the door for new investigations.

3.3.3 Working with a conceptual space

More generally, using the conceptual space might have various advantages and an important pragmatic aspect in that it can solve current problems in the literature and it can be used to test hypotheses and advance inquiries. It also opens the door to new kinds of inquiries. I want to elaborate a bit on this idea.

As mentioned earlier, conceptual spaces have a variety of interesting properties and manipulating them can bring us to see the processes under a new perspective. Not only would

---

[56] It may be possible that the reverse path happens, that a Type 1 process becomes Type 2, but in the vast majority of the cases discussed in the literature such a control of automatic processes is only possible indirectly (e.g. creating the appropriate setting to trigger given Type 1 processes; cf. Poirier and Beaulac, in preparation). After a period during which there is no practice, it might be harder to use a process that had been automatized, but I would not say the process has returned to be 'Type 2'.

[57] And it would most probably be *ad hoc* since, as I mentioned earlier, there is nothing in their framework allowing such modifications / transitions in the classification of a given process.

it be possible to identify clusters in this conceptual space, but it would also be possible to discover new clusters (or even interesting 'subclusters,' viz. groups of processes within a cluster that has interesting properties for a given theory or explanation) we currently have no way of predicting. For example, if a distribution of processes looks like what is illustrated in Figure 3.2, we might find useful and interesting to refer to the processes in the grey zone with a different label – or modify the way we talk about some of these 'in between' processes: for instance, the Type 3 Evans predicts we need in order to explain preattentive processes could be distinguished from a process that is only partly automatized. We would also understand, at least partly, the differences between these processes in the grey zone as we would have a precise characterization of their peculiarities, giving us an idea of why they cannot be explained with other accounts of dual-process theories.

Another possibility brought by this conceptual space framework is its manipulation: modifying it in various ways such as adding new dimensions or removing existing ones might reveal previously unknown features. The resulting conceptual space might be very different from the initial one; new clusters might appear, some might disappear, two apparently very different processes might be distinguished more sharply, and so on. The conceptual space can, in this manner, be a tool to test new hypotheses, to compare processes, to do research and try to explain and understand the mind. With this powerful tool that keeps the specificities of the processes it classifies, we might be able to take our understanding to a new level.

Resolving the unity problem is a concrete example where the manipulation of the conceptual space *could* be useful. With current tools, it is hard to identify the unifying characteristic of a given 'Type;' however, by removing some dimensions of the conceptual space and verifying if a given cluster is still to be found we have a first tool to identify these crucial characteristics. A second way to discriminate the weight of the characteristics would be to identify which are the characteristics that co-vary most strongly and find correlations with other fields of inquiry (e.g. neuroscience). This might not completely solve the unity problem – I only offer a tool that can help with this goal –, but I believe it is an interesting path to follow.

These remarks are, of course, speculative. However, none of the advantages I identified here is too far fetch: until now we did not have a similar way to map our knowledge of cognitive processes. Only using and elaborating this space will allow us to see where it can bring future research on the various dual-process theories currently in the literature (cf. Table 2.1).

3.4 Conclusion

In this chapter, I argued that Samuels' (2009) account of dual-process theories, while very interesting, has problems very similar to those of the other dual-process theories, discussed in Chapter II. However, the rejection of Samuels' account does not amount to a rejection of dual-process theories: I suggested that we could understand the Type 1 / Type 2 distinction as a continuum (Hammond, 1996), before putting forward reasons to doubt the adequacy of this idea. One continuum is not enough to take into account the complexity of cognitive processes, and we must therefore use many continua that will form an *n*-dimensional conceptual space where we will be able to place the processes and, eventually (hopefully!), identify clusters. Much of this discussion is speculative as this framework is currently not implemented in research, but I argued for several advantages of using such a powerful tool, such as its ability to model processes having crossover characteristics and the basis it offers to discuss how some processes change from Type 2 to Type 1.

Whether the difference between Type 1 and Type 2 is 'deep functional' or not could most probably be resolved depending on how the processes are clustering in the conceptual space. It is also possible that we will find different types of Type 1 processes, for example Fodor-modules and expert abilities. The only certitude that remains is the phenomenological difference Carruthers identified: some cognitive processes feel very different from others – although we could also imagine that two different Type 1 processes can feel different from one another. A lot of research still has to be done if we want to have a better understanding of the important difference identified by dual-process theories.

In the worst case scenario, the distinction between Type 1 and Type 2 processes is an interesting heuristic model that allows us to understand better some functional differences between processes (more or less what Carruthers, 2009 is suggesting). However, in the best of situations (the best for current dual-process theories, at the very least), the perspective offered by the conceptual space framework will be a key to illustrate and understand the duality of the mind that has been put forward by the dual-process theories. It will also be possible to understand and explain the processes that are not part of either Type, without 'forcing' them into a category in which they do not belong.

CONCLUSION

As mentioned at the beginning of Chapter II, dual-process theories are widely discussed and many researchers adopt a framework along the lines of these theories. There is no consensus however on the form these 'dual processes' should take, or which version of the theory is the most interesting. Some of these frameworks attracted a lot of attention and had quite an influence, such as Samuels' (2009) suggestion to divide the mind into two 'Types of processes,' instead of using the usual two 'Systems' as advocated in many of the initial dual-process / dual-system frameworks. The influence of Samuels' account is not surprising: his two way collapse, the collapse of both S1 and S2 into 'Types,' is an interesting solution to many problems of dual-process theories but, as I argued in 3.1, it is not enough to 'save' them. This is why I elaborated and suggested the conceptual space account in 3.2, as both a solution to the quibbles about 'module' and those about S1 / S2 or Type 1 / Type 2. This framework allows for a new examination of cognitive processes, and it contributes to a new way of defining the important notions of the field.

My take-home message would go like this: dual-process theories are very useful and interesting frameworks that provide us with explanations for many cognitive phenomena. Yet, they have important limitations and they cannot explain all of cognition: this is why we need to integrate these theories into a more comprehensive framework, like the one I tried to create in 3.2. This tool has the potential of offering interesting insights into the peculiarities of the mind.

Before arriving at this conclusion however, my argument had two major steps: I first discussed what 'module' means and, then, I examined what exactly the dual-process theorists are claiming, before putting the accent on the main shortcoming of their theories: the inability to offer a framework explaining all cognitive processes.

In Chapter I, I explored and tried to make sense of the debates around what 'module' means, or rather what it should mean. After a presentation of Fodor's very own 'Type 1,' Fodor-modules, and a discussion of how Evolutionary Psychologists began to also use the same notion, I argued there were many unresolved issues in both of these accounts, some of which Carruthers (2006) tries to address. I then described and criticized Carruthers' strong central modularity account, and I defended Samuels' (2006) more plausible weak central modularity hypothesis. This account, as I showed, is more or less another way to describe the mind as having dual processes, viz. the modular parts that are working in parallel and are fast, automatic, unconscious, etc. and central cognition, including some modular parts[58], is described as slow, serial, controlled, conscious, etc. This is not unlike Fodor's (1983) proposal, and I believe dual-process theories can be a wider framework to think about modules. In fact, even for a massive modularist like Carruthers (2006; 2009), while there is for him no deep functional difference between S1 and S2, recognizing and using the notion of S2 to explain cognitive phenomena can be very useful, and it allows for the distinction we intuitively make between two types of processing since they, at least, feel very different.

Chapter II was the ground for my critique of some of the proposed dual-process theories. I first explored what this family of theories is, before examining four important theories that are associated with this trend: Carruthers' (2009), Lieberman's (2007; 2009; Satpute and Lieberman, 2006), Stanovich's (especially his 2009 account) and Evans' (2009). By trying to explain cognition with a clear distinction between (only) two systems (or sets of systems)[59], every model examined in Chapter II (and Samuels', 2009 in 3.1) encounters some problems because at least one cognitive process, or one category of cognitive processes (e.g. automatized Type 2 processes) does not fit in each of the models. The list of characteristics that the processes (or these 'Systems') are thought to have in these accounts is

---

[58] These 'modular parts' have some of the characteristics Fodor would have attributed to 'central cognition.'

[59] Carruthers' account is, as we have seen, a bit different: for him, there really is just modules, and nothing else. As a reminder: S2 is realized in multiple iterations of modules (S1). Nevertheless, his account is not immune to my critique: some cognitive processes do not fit in his model.

just too rigid as it does not allow the explanation of processes having crossover characteristics.

Finally, in Chapter III, as mentioned above, I criticized Samuels' (2009) account, developed my proposal of a conceptual space and highlighted some of its advantages over other frameworks in the literature. While elaborating the conceptual space account, I examined the Type 1 / Type 2 distinction (the most plausible form out of the different dual-process theories I examined here) under a new light.

In the end, my current hypothesis is that cognitive scientists identified an important distinction under the labels 'Type 1' and 'Type 2,' a distinction that will most probably be present in the conceptual space. However, I also believe there will be other distinctions to be made, viz. I believe that other clusters will appear in the conceptual space. As mentioned in 3.3.2, one of the most obvious partitioning in my mind is the difference between 'natural' (evolved and adapted) and acquired Type 1 processes. There are most probably distinctions to be made in Type 2 processes as well (e.g. Stanovich's distinction between the algorithmic and reflective minds), and it would not be too surprising to find some clusters we cannot even predict because of our current ignorance of the relevant data or ill-advised efforts to fit data in current distinctions.

For now however, and this might be seen as a very controversial claim, the Type 1 / Type 2 distinction can only be taken as a heuristic distinction, and it should be considered as such by researchers – hence my radical subtitle, "for a heuristic interpretation of dual-process theories." There is simply not enough data (and there are some problems such as those I identified in this dissertation; other researchers, such as Keren and Schul, 2009, also criticize the dual-process framework) to support the claims that dual-process theories (as they are currently understood) can give us a *complete* understanding of the mind. Type 1 and Type 2 processes are not natural kinds as they are defined in the current state of dual-process theories (cf. Afterword), but they remain tremendously useful in organizing thinking in some fields of inquiry, such as in the psychology of reasoning. Yet, even if the distinction is just heuristic, I maintain it still identifies real differences between types of processes, but the available theories do not render these differences in an appropriate way. There is just no way,

in the current state of our knowledge about cognitive processes, to claim that Type 1 / Type 2 is a natural kind distinction and I hope that the conceptual space I defend provides tools to prove otherwise.

AFTERWORD

WHY SHOULD WE CARE ABOUT DUAL-PROCESS THEORIES?

As I mentioned more than once, my critique of many of the accounts we find in the dual-process theory literature is not intended to promote the idea that dual-process theories should be abandoned: *au contraire*, they should be pursued even further, but in an appropriate framework. I cannot stress this out too much: *there are processes that are dual*, and the evidence is overwhelming (Evans, 2008 offers a review of the literature). There is just no theory that currently succeeds by using *only* these dual processes to explain every cognitive process there is, and I believe it is hopeless to try elaborating such an architecture of the mind: this is an impossible mission, a point I made clearly in Chapter II and in 3.1. In short, there are dual-processes to be sure, but we have as of yet no adequate theory of dual-processes. We need a larger, more encompassing framework, like the one I developed in 3.2.

This important caveat aside, even current dual-process theories offer a very rich and interesting framework to tackle old problems in cognitive science and also in philosophy. In this afterword, I want to sketch what dual-process theories imply for diverse issues in philosophy, mostly for teaching critical thinking, but also for how we should think about knowledge in a naturalized framework. Dual-process theories, I believe, offer numerous prospects for future research, including but not limited to: pedagogical recommendations about teaching critical thinking and logic (Beaulac and Robert, submitted; Stanovich and West, 2008), and reflection on the building of 'cognitive niches' to enhance cognition (Poirier and Beaulac, in preparation). Also, such a framework can bring new ideas about the

nature of knowledge, and it can push further the research on knowledge as a natural kind, building on work such as Clarke's (2004) work in epistemology (Beaulac, in preparation; Clarke, in preparation). Recent proposals also argue dual-process theories might even provide us with a framework to understand concepts in psychology (Clarke, in preparation; Piccinini, forthcoming; Poirier and Beaulac, forthcoming; but see Machery, forthcoming for a reply in which we can find objections to dual-process theories discussed in this dissertation).

This work is only a small part of where a dual-process perspective on the mind can take us. To give a few other examples (I have no pretention of being exhaustive here): Darlow and Sloman, 2010 and de Sousa (2010) have applied similar accounts to the understanding of emotions; much work is being (and has been) done in moral psychology and in ethics (e.g. Cushman, Young and Greene, 2010; Haidt, 2001); insights from dual-process theories can inform work in logic and in the psychology of reasoning (e.g. Evans and Over, 2004; also Mercier and Sperber, forthcoming for an account of the evolution of reasoning using a dual-process framework); it can also provide a framework for accounts of judgment and decision making (Sahlin, Wallin and Persson, 2010; also Bishop and Trout, 2005 for a framework inspired from the heuristics and biases literature). There are also many fields of philosophical inquiry, where I know of no work that has been done for the moment, such as free will and the role and nature of self control mechanisms (cf. Hassin, Ochsner and Trope, 2010) or research in the philosophy of language (building on work such as Stainton's 2006a work on pragmatics) that could benefit from the insights of dual-process theories. Moreover, some recent proposals also have strong similarities with dual-process theories: I have in mind here work such as Szabó-Gendler's (2008a; 2008b) on the notions of 'alief' and 'belief,' where alief can easily be described as the 'Type 1' counterpart of belief.

5.1 Dual-Process Epistemology?

Type 1 processes sometimes go against our best interest: our intuitions sometimes lead us astray. Yet, eliminating the use of Type 1 processing entirely (as suggested by Bishop and Trout, 2005) is both unpractical and problematic on a number of levels. First, we cannot eliminate these *automatic* processes completely, only learn to circumvent them. Second, we might be able to replace, institutionally, human judgment with decision-making algorithms, but these rules cannot account for the rare exceptions that would be detected by specialists (e.g. Dreyfus and Dreyfus', 1986 discussions of the work of nurses). It is important to note, however, that these algorithms often produce statistically better results (cf. Gigerenzer, 2007, chapter 9). The third problem is that, in our everyday lives, we are faced with too complex and too diversified situations to create such rules for every situation. In most situations, our intuitions actually are more efficient and effective (cf. Gigerenzer *et al.*, 1999; Gigerenzer, 2007).

Yet, even if the human mind is a kluge fraught with cognitive biases and limitations, we are not powerless in front of its 'weaknesses.' We are not, in any way, condemned by them, unless we decide to ignore them. We have, after all, discovered many of the mind's shortcomings – although probably not all of them, and we might be far from having done so. Just as in other areas of human endeavor, discovering a flaw in a process constitutes an invitation to correct it, by changing the process if we can, or by writing or building a 'patch' when we cannot: this is what software engineers do each time they discover a flaw in a large program; this is also what NASA did when it discovered that the Hubble's flawed central mirror made it myopic. In the case of minds, we have also developed such 'patches,' which are called mindwares in the literature, viz. "whatever people can learn that helps them to solve problems, make decisions, understand difficult concepts, and perform other intellectually demanding tasks better" (Perkins, 1995, 13). As Gigerenzer puts it: "[...] the real question is not *if* but *when* can we trust our guts? To find the answer, we must figure out how intuition actually works in the first place." (2007, 17) And when we know that a particular intuition is not reliable, we can devise mindwares that work around these biases and limitations, helping us reason better, going against this intuition in some contexts and –

more generally – abling us to see if our current reasoning practices are, or not, an optimal way to confront a given challenge.

Stanovich and West's (2008) model of reasoning allows for an even better understanding of why (effectively) teaching critical thinking is hard. They developed the idea that possessing the right mindware is not enough to avoid biases and errors coming from our intuitions. In epistemological terms, they are not enough to increase the reliability of our Type 1 mechanisms. Of course, what they call the *mindware gap* must be filled by learning the appropriate mindware, but, as Kahneman and Frederick rightly pointed out, something else is needed: one must "possess the relevant logical rules and also to recognize the applicability of these rules in particular situations" (2002, 68). In order to successfully use a given mindware, there must be override detection, that is, the ability to detect that the mindware in question should be used in a particular situation. Thus, having the mindware – the *algorithm* (remember Stanovich's, 2009 subdivision of S2, cf. footnote 31) – is not enough. Several studies support this claim (e.g. Houdé and Moutier, 1996; 1999) and Houdé *et al.*'s (2000) study is a very good example of how we can successfully override the Type 1 processes. It follows that possessing the right mindwares makes it possible to overcome Type 1 processes' limitations (Stanovich, 2004; Stanovich *et al.*, 2008) only if knowledge about overriding is also provided (Houdé *et al.*, 2000; Stanovich, 2009; Stanovich and West, 2008)[60].

I believe that integrating mindwares into our account of knowledge is essential in order to make room for all that matters concerning knowledge[61], since we have the ability to correct

---

[60] This may mean there are different types and different levels of mindwares, but such a claim goes beyond the scope of what I want to discuss here. Thanks to Jean-Pierre Marquis for pointing this out.

[61] And, perhaps, not exclusively for human knowledge. There is evidence that certain animals do have processes exhibiting a number of Type 2 properties; therefore these animals could have Type 2 knowledge, at least to a certain degree (e.g. they might not have knowledge about the limitations of their Type 1 processes, but they clearly have the capacity for action-rehearsal as discussed by Carruthers, 2006).

the outputs of our generally reliable knowledge mechanisms (Clarke's, 2004 modular knowledge) when we determine that they are less than reliable in certain contexts. What I suggest for further research centered on debates in epistemology is that a part of what is understood as knowledge depends on mindwares – the cultural heuristics – developed to correct biases inherent to some modules. For the belief that the two lines are the same length in the Müller-Lyer illusion to be considered knowledge, we need an account of knowledge that includes something along the lines of the capacity to elaborate rules that compel the use of a ruler each time there is the need to determine the length of a line in a tricky epistemic context. This epistemic context can be identified because there is (scientific) knowledge regarding the limitations inherent to humans' evolved cognition; in this case, we know that the human visual system can be subject to such illusions. In other words, in the Müller-Lyer case, knowledge that the two lines are of the same length is linked to the knowledge of the unreliability of vision in this context, and to the knowledge that a ruler is a reliable tool in order to obtain this knowledge.

It is thus possible, I believe, to devise a view of knowledge that is closer to dual-process theories. According to this view, there are two types of knowledge: modular knowledge that results from modules working autonomously (arguably identified by externalist accounts of knowledge) and reflective knowledge that results from our Type 2 processes, which are not cognitively closed, and can thus acquire mindwares and be influenced by culture (arguably what internalists identify). While we have no control over the first type of knowledge – e.g. it is impossible not to see when our eyes are open – we can control, and acquire new mindwares to control, Type 2 knowledge (cf. Beaulac, in preparation).

From this perspective, we acquire culturally-designed tools to augment the reliability of modular knowledge, and thus learn to go beyond the Type 1 processes' suboptimal responses to some complex problems. Understanding the mind within the dual-process framework, individual learning, *qua* acquiring the means to have Type 2 knowledge, can be seen as the cultural diffusion of a collection of cultural items (tools, heuristics). These will enhance the response of Type 2 processes that need to be activated and must have the correct tools to produce a correct answer (cf. Poirier and Beaulac, in preparation). In epistemology, this means that the reliability of modular knowledge can, with the use of mindwares, be

increased. This view of knowledge is, I believe, more comprehensive than most other naturalized accounts of knowledge.

My proposition then is to be understood along these lines: a more inclusive conception of knowledge would have to include the methods of science in order (i) to have a better understanding of the cognitive structures underlying knowledge, which includes learning about their limitations, (ii) to develop that tools and heuristics necessary to correct, or ameliorate, human cognition – what Clarke (2004) calls the *meliorative project*[62] and (iii) to make easier the recognition of the epistemic limitations of Type 1 processes in order to be able to identify when it is necessary to use Type 2 processes. As Clarke explains, we must "[f]irst, work at understanding early-developing, task-specific mechanisms and; second, study how further cognitive processes are built on top of the initial ones to overcome their limitations" (2004, 40). I hope that the material I developed in this dissertation can help advancing both of these goals.

---

[62] Bishop and Trout (2005) call this Ameliorative Psychology.

REFERENCES

Atkinson, A.P. and M. Wheeler. 2004. "The Grain of Domains: The Evolutionary-Psychological Case Against Domain-General Cognition". *Mind and Language*, vol. 19, n° 2, pp. 147-176.

Baars, B.J. 1988. *A Cognitive Theory of Conciousness*. New York: Cambridge University Press.

Baars, B.J. 1997. *In the Theater of Conciousness. The Workspace of the Mind.* New York: Oxford University Press.

Bailin, S. and H. Siegel. 2003. "Critical Thinking". In *The Blackwell Guide to the Philosophy of Education* edited by N. Blake, P. Smeyers, R. Smith and P. Standish, pp. 181-193.

Baillargeon, N. 2005. *Petit Cours d'Autodéfense Intellectuelle*. Montréal: Lux Éditeur.

Barkow, J.H., L. Cosmides and J. Tooby, eds. 1992. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.

Barrett, H.C. 2009. "Les modules "en chair et en os"". In *Les mondes darwiniens. L'évolution de l'évolution* edited by T. Heams, P. Huneman, G. Lecointre and M. Silberstein, pp. 779-786.

Barrett, H.C. and R. Kurzban. 2006. "Modularity in Cognition: Framing the Debate". *Psychological Review*, vol. 113, n° 3, pp. 628-647.

Beaulac, G. In preparation. "Dual-Process Epistemology. Characterizing Knowledge as Modular *and* Reflective".

Beaulac, G. and S. Robert. Submitted to *Les ateliers de l'éthique*. "Les théories de l'éducation à l'ère des sciences cognitives: le cas de l'enseignement de la pensée critique et de la logique".

Bechtel, W. and A. Abrahamsen. 2001. *Connecionism and the Mind. Parallel Processing, Dynamics, and Evolution in Networks*. New York: Blackwell.

Bickle, J., P. Mandik and A. Landreth. 2010. "The Philosophy of Neuroscience". *Stanford Encyclopedia of Philosophy*, online: <http://plato.stanford.edu/entries/neuroscience/>.

Bishop, M. and J.D. Trout. 2005. *Epistemology and the Psychology of Human Judgment*. New York: Oxford University Press.

Boyd, R. 1990. "What realism implies and what it does not". *Dialectica*, vol. 43, pp. 5-29.

Boyd, R. 1991. "Realism, anti-foundationalism and the enthusiasm for natural kinds". *Philosophical Studies*, vol. 61, pp. 127-148.

Breiman, L., J. Friedman, R.A. Olshen and C.J. Stone. 1984. *Classification and Regression Trees*. Monterey (CA): Wadsworth & Brooks.

Buller, D.J. 2005. *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*. Cambridge (MA): MIT Press.

Carruthers, P. 2004. "Practical Reasoning in a Modular Mind". *Mind and Language*, vol. 19, n° 3, pp. 259-278.

Carruthers, P. 2006. *The Architecture of Mind*. New York: Oxford University Press.

Carruthers, P. 2009. "An architecture for dual reasoning". In Evans and Frankish 2009, pp. 109-128.

Chaiken, S. 1980. "Heuristic versus systematic information processing and the use of source versus message cues in persuasion". *Journal of Personality and Social Psychology*, vol. 39, pp. 752-766.

Chen, S. and S. Chaiken. 1999. "The heuristic-systematic model in its broader context". In *Dual-process theories in social psychology* edited by S. Chaiken and Y. Trope, pp. 73-96.

Churchland, P. 1989. *A Neurocomputational Perspective. The Nature of Mind and the Structure of Science*. Cambridge (MA): MIT Press.

Clarke, M. 2004. *Reconstructing Reason and Representation*. Cambridge (MA): MIT Press.

Clarke, M. In preparation. "Concepts, Intuitions, and Epistemic Norms".

Cohen, J.R. and M.D. Lieberman. 2010. "The Common Neural Basis of Exerting Self-Control in Multiple Domains". In Hassin, Ochsner and Trope 2010, pp. 141-160.

Cosmides, L. and J. Tooby. 1987. "From evolution to behavior. Evolutionary psychology as the missing link". In *The latest on the best: Essays on evolution and optimality* edited by J. Dupré, pp. 277-306.

Cosmides, L. and J. Tooby. 1994. "Origins of domain-specificity: The evolution of functional organization". In Hirschfeld and Gelman 1994, pp. 85-116.

Cosmides, L. and J. Tooby. 2000. *Evolutionary Psychology : A Primer*, online: <http://www.psych.ucsb.edu/research/cep/primer.html>.

Cushman, F., L. Young and J.D. Greene. 2010. "Our multi-system moral psychology: Towards a consensus view". In *The Moral Psychology Handbook* edited by J.M. Doris and the Moral Psychology Research Group, pp. 47-71.

Darlow, A.L. and S.A. Sloman. 2010. "Two systems of reasoning: architecture and relation to emotion". *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 1, n° 1, pp. 1-11.

Dawkins, R. 1986. *The Blind Watchmaker. Why the Evidence of Evolution Reveals a Universe Without Design*. New York: Norton.

de Sousa, R. 2010. "The Mind's Bermuda Triangle: Philosophy of Emotions and Empirical Science". In *The Oxford Handbook of Philosophy of Emotion* edited by P. Goldie, pp. 95-117.

Downes, S.M. 2008. "Evolutionary Psychology". *Stanford Encyclopedia of Philosophy*, online: <http://plato.stanford.edu/entries/evolutionary-psychology/>.

Dreyfus, H.L. and S.E. Dreyfus. 1986. *Mind Over Machine. The Power of Human Intuition and Expertise in the Era of the Computer*. New York: The Free Press.

Eliasmith, C. and C.H. Anderson. 2002. *Neural Engineering. Computation, Representation, and Dynamics in Neurobiological Systems*. Cambridge (MA): MIT Press.

Epstein, S. 1994. "Integration of the cognitive and psychodynamic unconscious". *American Psychologist*, vol. 49, pp. 709-724.

Epstein, S. and R. Pacini. 1999. "Some basic issues regarding dual-process theories from the perspective of cognitive-experiential theory". In *Dual-process theories in social psychology* edited by S. Chaiken and Y. Trope, pp. 462-482.

Ermer, E., L. Cosmides and J. Tooby. 2007. "Functional Specialization and the Adaptationist Program". In *The Evolution of Mind. Fundamental Questions and Controversies* edited by S.W. Gangestad and J.A. Simpson, pp. 153-160.

Evans, J.St.B.T. 1989. *Bias in Human Reasoning: Causes and Consequences*. Brighton: Erlbaum.

Evans, J.St.B.T. 1998. "Matching bias in conditional reasoning: do we understand it after 25 years?". *Thinking and Reasoning*, vol. 4, pp. 45-82.

Evans, J.St.B.T. 2003. "In two minds: dual-process accounts of reasoning". *TRENDS in Cognitive Sciences*, vol. 7, n° 10, pp. 454-459.

Evans, J.St.B.T. 2006. "The heuristic-analytic theory of reasoning: Extension and evaluation". *Psychonomic Bulletin & Review*, vol. 13, n° 3, pp. 378-395.

Evans, J.St.B.T. 2007. "On the resolution of conflict in dual-process theories of reasoning". *Thinking and Reasoning*, vol. 13, pp. 321-329.

Evans, J.St.B.T. 2008. "Dual-Processing Accounts of Reasoning, Judgment and Social Cognition". *Annual Review of Psychology*, vol. 59, pp. 255-278.

Evans, J.St.B.T. 2009. "How many dual-process theories do we need? One, two, or many?". In Evans and Frankish 2009, pp. 33-54.

Evans, J.St.B.T. and K. Frankish, eds. 2009. *In Two Minds. Dual Processes and Beyond.* New York: Oxford University Press.

Evans, J.St.B.T. and D.E. Over. 1996. *Rationality and Reasoning*. Hove (UK): Psychology Press.

Evans, J.St.B.T. and D.E. Over. 2004. *If.* New York: Oxford University Press.

Faucher, L. and P. Poirier. 2009. "Philosophie: Modularité et psychologie évolutionniste". In *Darwin en tête! L'Évolution et les sciences cognitives* edited by J.-B. Van der Henst and H. Mercier, pp. 275-308.

Fitch, W.T. 2010. *The Evolution of Language*. New York: Cambridge University Press.

Fitts, P.M. and M.I. Posner. 1967. *Human Performance*. Belmont (CA): Brooks / Cole Publishing.

Fodor, J. 1983. *The Modularity of Mind*. Cambridge (MA): MIT Press.

Fodor, J. 1998. *Concepts: Where Cognitive Science Went Wrong*. New York: Oxford University Press.

Fodor, J. 2000. *The Mind Doesn't Work That Way*. Cambridge (MA): MIT Press.

Fodor, J. and M. Piattelli-Palmarini 2010. *What Darwin Got Wrong*. New York: Farrar, Straus and Giroux.

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

Frankish, K. and J.St.B.T. Evans. 2009. "The duality of mind: An historical perspective". In Evans and Frankish 2009, pp. 1-29.

Gärdenfors, P. 2000. *Conceptual Spaces: The Geometry of Thought*. Cambridge (MA): MIT Press.

Gigerenzer, G. 2007. *Gut Feelings. The Intelligence of the Unconsciours*. New York: Penguin Books.

Gigerenzer, G., J. Czerlinski and L. Martignon. 2002. "How Good Are Fast and Frugal Heuristics?". In Gilovich, Griffin and Kahneman 2002, pp. 559-581.

Gigerenzer, G., P.M. Todd and ABC Research Group. 1999. *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Gilovich, T., D. Griffin and D. Kahneman, eds. 2002. *Heuristics and Biases. The Psychology of Intuitive Judgment*. New York: Cambridge University Press.

Goel, V. 2007. "Anatomy of deductive reasoning". *TRENDS in Cognitive Sciences*, vol. 11, n° 10, pp. 435-441.

Goldman, A.I. 1991. *Liaisons: Philosophy Meets the Cognitive and Social Sciences*. Cambridge (MA): MIT Press.

Gopnik, A.M. and A.N. Meltzoff. 1996. *Words, Thoughts and Theories*. Cambridge (MA): MIT Press.

Gopnik, A.M., A.N. Meltzoff and P.K. Kuhl. 2000. *The Scientist in the Crib: What Early Learning Tells Us About the Mind*. New York: HarperCollins.

Haidt, J. 2001. "The emotional dog and its rational tail: A social intuitionist approach to moral judgment". *Psychological Review*, vol. 108, pp. 814-834.

Hammond, K.R. 1996. *Human judgment and social policy*. New York: Oxford University Press.

Hassin, R.R., K.N. Ochsner and Y. Trope, eds. 2010. *Self Control in Society, Mind, and Brain*. New York: Oxford University Press.

Hauser. M.D. 2000. *Wild Minds: What Animals Really Think*. New York: Henry Holt.

Hirschfeld, L. and S. Gelman, eds. 1994. *Mapping the Mind: Domain-specificity in cognition and culture*. New York: Cambridge University Press.

Holt, J. "Blindsight in Debates About Qualia". *Journal of Consciousness Studies*, vol. 6, n° 5, pp. 54-71.

Houdé, O. and S. Moutier. 1996. "Deductive Reasoning and Experimental Inhibition Training: The Case of the Matching Bias". *Current Psychology of Cognition*, vol. 15, pp. 409-434.

Houdé, O. and S. Moutier. 1999. "Deductive Reasoning and Experimental Inhibition Training: The Case of the Matching Bias. New Data and Reply to Girotto". *Current Psychology of Cognition*, vol. 18, pp. 75-85.

Houdé, O., L. Zago, E. Mellet, S. Moutier, A. Pineau, B. Mazoyer and N. Tzourio-Mazoyer. 2000. "Shifting from the Perceptual Brain to the Logical Brain: The Neural Impact of Cognitive Inhibition Training". *Journal of Cognitive Neuroscience*, vol. 12, n° 5, pp. 721-728.

Jacob, F. 1970. *La logique du vivant, une histoire de l'hérédité*. Paris: Gallimard.

Kahneman, D. and S. Frederick. 2002. "Representativeness revisited: Attribute substitution in intuitive judgment". In Gilovich, Griffin and Kahneman 2002, pp. 49-81.

Kahneman, D. and S. Frederick. 2005. "A Model of Heuristic Judgment". In *The Cambridge Handbook of Thinking and Reasoning* edited by K.J. Holyoak and R.G. Morrison, pp. 267-293.

Karmiloff-Smith, A. 1992. *Beyond Modularity. A Developmental Perspective on Cognitive Science*. Cambridge (MA): MIT Press.

Keren, G. and Y. Schul. 2009. "Two Is Not Always Better Than One. A Critical Evaluation of Two-System Theories". *Perspectives on Psychological Science*, vol. 4, n° 6, pp. 533-550.

Kirk, R. 1974. "Sentience and Behaviour". *Mind*, vol. 83, pp. 43-60.

Krug, M.K. and C.S. Carter. 2010. "Anterior Cingulate Cortex Contributions to Cognitive and Emotional Processing: A General Purpose Mechanism for Cognitive Control and Self-Control". In Hassin, Ochsner and Trope 2010, pp. 3-26.

Leslie, A.M. 1994. "ToMM, ToBY, and Agency: Core architecture and domain specificity". In Hirschfeld and Gelman 1994, pp. 119-148.

Lieberman, M.D. 2003. "Reflective and reflexive judgment processes: a social cognitive neuroscience approach". In *Social Judgments: Implicit and Explicit Processes* edited by J.P. Forgas, K.D. Williams and W. Von Hippel, pp. 44-67.

Lieberman, M.D. 2007. "Social Cognitive Neuroscience: A Review of Core Processes". *Annual Review of Psychology*, vol. 58, pp. 259-289.

Lieberman, M.D. 2009. "What zombies can't do: A social cognitive neuroscience approach to the irreducibility of reflective consciousness". In Evans and Frankish 2009, pp. 293-316.

Machery, E. 2009. *Doing Without Concepts*. New York: Oxford University Press.

Machery, E. Forthcoming. "Replies to Lombrozo, Piccinini, and Poirier and Beaulac". *Dialogue*.

Marcus, G.F. 2004. *The Birth of Mind: How a Tiny Number of Genes Creates the Complexity of Human Thought*. New York: Basic Books.

Marcus, G.F. 2008. *Kluge. The Haphazard Construction of the Human Mind.* New York: Houghton Mifflin Harcourt.

McGurk, H. and J. Macdonald. 1976. "Hearing lips and seeing voices". *Nature*, vol. 264, n° 5588, pp. 746-748.

Mercier, H. and D. Sperber. Forthcoming. "Why do humans reason? Arguments for an argumentative theory". *Behavioral and Brain Sciences*. Available online: <http://www.dan.sperber.fr/wp-content/uploads/2009/10/MercierSperberWhydohumansreason.pdf>.

Moshman, D. 2000. "Diversity in reasoning and rationality: Metacognitive and developmental considerations". *Behavioral and Brain Scineces*, vol. 23, pp. 689-690.

Newell, A. and H.A. Simon. 1976. "Computer Science as Empirical Enquiry: Symbols and Search". *Communications of the ACM*, vol. 19, n° 3, pp. 113-126.

Nisbett, R., K. Peng, I. Choi and A. Norenzayan. 2001. "Culture and systems of thought: Holistic vs analytic cognition". *Psychological Review*, vol. 108, pp. 291-310.

Osman, M. 2004. "An evaluation of dual-process theories of reasoning". *Psychonomic Bulletin & Review*, vol. 11, n° 6, pp. 988-1010.

Perkins, D. 1995. *Outsmarting IQ: The Emerging Science of Learnable Intelligence*. New York: Free Press.

Piccinini, G. Forthcoming. "Two Kinds of Concept: Implicit and Explicit". *Dialogue*.

Pinker, S. 1997. *How the Mind Works*. New York: Norton.

Poirier, P. In preparation. "Can Blackboards Save Classical Computation?".

Poirier, P. and G. Beaulac. Forthcoming. "Le véritable retour des définitions". *Dialogue*.

Poirier, P. and G. Beaulac. In preparation. "La construction d'algorithmes et de niches cognitives".

Pritchard, D. 2005. "Virtue Epistemology and the Acquisition of Knowledge". *Philosophical Explorations*, vol. 8, n° 3, pp. 229-243.

Prinz, J. 2006. "Is the Mind Really Modular?". In Stainton 2006b, pp. 22-36.

Reber, A.S. 1993. *Implicit Learning and Tacit Knowledge*. Oxford: Oxford University Press.

Rosenbaum, D.A., J.S. Augustyn, R.G. Cohen and S.A. Jax. 2006. "Perceptual-Motor Expertise". In *The Cambridge Handbook of Expertise and Expert Performance* edited by K.A. Ericsson, N. Charness, P.J. Feltovich and R.R. Hoffman, pp. 505-520.

Sahlin, N.-E., A. Wallin and J. Persson. 2010. "Decision science: from Ramsey to dual process theories". *Synthese*, vol. 172, pp. 129-143.

Samuels, R. 1998. "Evolutionary Psychology and The Massive Modularity Hypothesis". *British Journal for the Philosophy of Science*, vol. 49, pp. 575-602.

Samuels, R. 2006. "Is the Human Mind Massively Modular?". In Stainton 2006b, pp. 37-56.

Samuels, R. 2009. "The magical number two, plus or minus: Dual-process theory as a theory of cognitive kinds". In Evans and Frankish 2009, pp. 129-146.

Samuels R., S. Stich and M. Bishop. 2002. "Ending the Rationality Wars: How To Make Disputes About Human Rationality Disappear". In *Common Sense, Reasoning and Rationality* edited by R. Elio, pp. 236-268.

Satpute, A.B. and M.D. Lieberman. 2006. "Integrating automatic and controlled processing into neurocognitive models of social cognition". *Brain Research*, vol. 1079, pp. 86-97.

Schneider, W. and R.M. Shiffrin. 1977. "Controlled and automatic human information processing I: Detection, search and attention". *Psychological Review*, vol. 84, pp. 1-66.

Scott, R.M. et R. Baillargeon. 2009. "Which penguin is this? Attributing false beliefs about object identity at 18 months". *Child development*, vol. 80, n° 4, pp. 1172-1196.

Simon, H.A. 1957. *Models of Man: Social and Rational*. New York: Wiley.

Sloman, S.A. 1996. "The empirical case for two systems of reasoning". *Psychological Bulletin*, vol. 119, pp. 3-22.

Sloman, S.A. 2002. "Two systems of reasoning". In Gilovich, Griffin and Kahneman 2002, pp. 379-398.

Smith, E.R. and J. DeCoster. 2000. "Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems". *Personality and Social Psychology Review*, vol. 4, n° 2, pp. 108-131.

Stainton, R.J. 2006a. *Words and Thoughts. Subsentences, Ellipsis, and the Philosophy of Language*. New York: Oxford University Press.

Stainton, R.J., ed. 2006b. *Contemporary Debates in Cognitive Science*. New York: Blackwell.

Stanovich, K.E. 1999. *Who is Rational? Studies of Individual Differences in Reasoning.* Mahway (NJ): Erlbaum.

Stanovich, K.E. 2004. *The Robot's Rebellion. Finding Meaning in the Age of Darwin.* Chicago: University of Chicago Press.

Stanovich, K.E. 2009. *What Intelligence Tests Miss: The Psychology of Rational Thought.* New Haven: Yale University Press.

Stanovich, K. E., M.E. Toplak and R.F. West. 2008. "The development of rational thought: A taxonomy of heuristics and biases". *Advances in Child Development and Behaviour*, vol. 36, pp. 251-285.

Stanovich, K.E. and R.F. West. 2000. "Individual differences in reasoning: Implications for the rationality debate?". *Behavioral and Brain Sciences*, vol. 23, n° 5, pp. 645-665.

Stanovich, K. E. and R.F. West. 2007. "Natural myside bias is independent of cognitive ability". *Thinking and Reasoning*, vol. 13, n° 3, pp. 225-247.

Stanovich, K.E. and R.F. West. 2008. "On the Relative Independence of Thinking Biases and Cognitive Ability". *Journal of Personality and Social Psychology*, vol. 94, n° 4, pp. 672-695.

Stenning, K. and M. van Lambalgen. 2008. *Human Reasoning and Cognitive Science.* Cambridge (MA): MIT Press.

Strack, F. and R. Deutsch. 2004. "Reflective and impulsive determinants of social behavior." *Personality and Social Psychology Review*, vol. 8, n° 3, pp. 220-247.

Szabó-Gendler, T. 2008a. "Alief and Belief". *Journal of Philosophy*, vol. 105, n° 10, pp. 634-663.

Szabó-Gendler, T. 2008b. "Alief in Action (and Reaction)". *Mind and Language*, vol. 23, n° 5, pp. 552-585.

Toates, F. 2006. "A model of the hierarchy of behaviour, cognition and consciousness". *Consciousness & Cognition*, vol. 15, pp. 75-118.

Todd, P. 2001. "Fast and frugal heuristics for environmentally bounded minds". In *Bounded Rationality: The Adaptive Toolbox* edited by G. Gigerenzer and R. Selten, pp. 51-70.

Tooby, J. and L. Cosmides. 1992. "The Psychological Foundations of Culture". In Barkow, Cosmides and Tooby, 1992, pp. 19-136.

Tversky, A. and D. Kahneman. 1974. "Judgement under uncertainty: heuristics and biases". *Science*, vol. 185, pp. 1124-1131.

Wason, P.C. and J.St.B.T. Evans. 1975. "Dual processes in reasoning?". *Cognition*, vol. 3, pp. 141-154.

West, R.F., M.E. Toplak and K.E. Stanovich. 2008. "Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions". *Journal of Educational Psychology*, vol. 100, pp. 930-941.

Wilson, T.D. 2002. *Strangers to Ourselves*. Cambridge (MA): Belknap Press.

Wimsatt, W.C. 1986. "Forms of Aggregativity". In *Human Nature and Natural Knowledge*, edited by A. Donagan, A.N. Perovich Jr. and M.V. Wedin, pp. 259-291.

Wimsatt, W.C. 2007. *Re-Engineering Philosophy For Limited Beings. Piecewise Approximations to Reality*. Cambridge (MA): Harvard University Press.

Woodward, J. and J. Allman. 2008. "Moral intuition: Its neural substrates and normative significance". *Journal of Physiology-Paris*, vol. 101, pp. 179-202.