HOW MANY MINDS DO WE NEED? TOWARD A ONE-SYSTEM ACCOUNT OF HUMAN REASONING

JOSHUA MUGG

A DISSERTATION SUBMITTED TO THE
FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN PHILOSOPHY
YORK UNIVERSITY
TORONTO, ONTARIO

MAY 2015

Abstract

To explain data from the reasoning and decision-making literature, dual-process theorists claim that human reasoning is divided: Type-1 processes are fast, automatic, associative, and evolutionarily old, while Type-2 processes are slow, effortful, rule-based, and evolutionarily new. Philosophers have used this distinction to their own philosophic ends in moral reasoning, epistemology, and philosophy of mind. I criticize dual-process theory on conceptual and empirical grounds and propose an alternative cognitive architecture for human reasoning.

In chapter 1, I identify and clarify the key elements of dual-process and dual-system theory. Then, in chapter 2, I undercut an inference to the best explanation for dual-process theory by offering a one-system alternative. I argue that a single reasoning system can accomplish the explanatory work done by positing two distinct processes or systems. In chapter 3, I argue that a one-system account of human reasoning is empirically testable—it is incompatible with there being contradictory beliefs that are produced by simultaneously occurring reasoning processes. I further argue, contra Sloman (1996), that we do not have evidence for such beliefs. Next, in chapter 4, I argue that the properties used to distinguish Type-1 from Type-2 processes cross-cut each other (e.g. there are evolutionarily new processes that are effortless). The upshot is that even if human reasoning were divided, it would not parse neatly into two tidy categories: 'Type-1' and 'Type-2.' Finally, in chapter 5, I fill in the details of my own one-system alternative. I argue that there is one reasoning system that can operate in many modes: consciously or unconsciously, automatically or controlled, and inductively or deductively. In contrast to the dual-process theorists, these properties do not cluster. For each property pair (e.g. automatic/controlled), and for a single instance of a task, the reasoning system will operate in a definitive mode. The reasoning system is like a mixing board: it has several switches and slides, one for each property pair. As subjects work through problems, they can alter the switches and slides—they can, perhaps unconsciously, change the process they use to complete the problem.

Dedication

To Patrick O’Neil Copley and James Edward Mugg. Grandad and Papa, respectively.

Acknowledgements

Thank you to Muhammad Ali Khalidi, Kristin Andrews, and Jacob Beck for their help and encouragement on this project. Muhammad Ali, or MAK, has been a wonderful supervisor. I frequently arrived at his office with a long list of questions, and always left with a clear direction. MAK not only made me a better philosopher, but also made me a better writer. Kristin deserves special thanks as well. Her piercing comments usually forced me into a new literature, and I am better for it. She was also a constant source of encouragement as I prepared portions of this dissertation for conferences and struck out on the job market. To all three of you, thank you.

Many thanks to the graduate students at York University for helpful conversations, and for looking over portions of this dissertation amended for conference presentations. Specifically, thank you Devin Curry, Eric Soler, Olivia Sultanescu, Ian Wright, and Marc Champagne. Most of the work in this dissertation was presented at conferences, and I would like to thank audience members and commentators at the European Society for Philosophy and Psychology, the Cognitive Science Society, the Society for Philosophy and Psychology, and the Canadian Philosophical Association. I would like to thank Steven Sloman for helpful comments on a very early draft of chapter 3. Finally, I would like to thank my wife, Sharon Mugg, for her stylistic suggestions on the prose, and for her patience with her philosopher husband.

This research was funded by a grant from the Social Sciences Research Council of Canada, and by a Provost Dissertation Scholarship from York University.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction, Motivation, and Background

How many minds do humans have? Recently a number of psychologists and philosophers have argued that humans have two minds. Furthermore, they are not talking about split-brain patients or subjects with psychological disorders; their claim is that ordinary humans have two minds. Call this claim the **two-mind theory**. Why would someone think the two-mind theory is true? One reason offered is that cognition is divided in two—that there are two systems that operate relatively independently of one another and are sometimes in conflict with one another. These two systems are named System 1 and System 2 (henceforth S1 and S2 respectively). Roughly, S1 is fast, evolutionarily old, associative or heuristically based, and automatic, whereas S2 is slow, evolutionarily new, rule-based, and controlled. In addition to the theory's prominence among psychologists as an explanation of human reasoning performance, some philosophers are now using this distinction to their own philosophical ends without critically examining the two-system theory itself. I will argue that the two-system and two-mind hypotheses face serious problems, and that humans possess only one reasoning system.

1. Background

Before going any further, it will be helpful to examine some specific two-system and two-mind theories. In doing so, a pattern will emerge. While there is some disagreement as to the relation of the two systems, the properties two-system and two-mind theorists use to distinguish the two systems tend to overlap. This overlap is commonly called the 'Standard Menu.' It is for this reason that these theories are appropriately categorized together as 'dual-system,' 'dual-process,' 'two-system,' or 'two-mind' theories. I begin by examining some of the most important psychologists and philosophers who advocate a two-system and (some) a two-mind theory.

1.1. Stanovich

It was Keith Stanovich who first introduced the terms 'System 1' and 'System 2' in 1999, though the concepts had existed for at least a decade before. Stanovich's primary reason for introducing his own two system theory is to find a way between two sides of a debate concerning

the degree of human rationality (or irrationality). On the one side are the Meliorists, who think that humans are, basically, irrational. The heuristics and biases literature (stemming largely from Kahneman and Tversky) has been dominated by Meliorists, and some philosophers, such as Stich (1990), fall into this camp as well. Stanovich himself tells us that his research comes out of the Meliorist tradition. However, there is another tradition (Stanovich calls it Panglossian) which claims that humans are basically rational. Stanovich (1999) cites Dennett and Davidson as his primary examples from philosophy, and (in his 2011) he cites evolutionary psychologists, adaptationist modelers, and ecological theorists as examples from psychology (such as Cosmides, Tooby, Gigerenzer, Oaksford, Chater, and Todd). According to Dennett and Davidson, widespread irrationality is impossible, since (to oversimplify) irrationality can only be understood against a background of overall rationality.[1] A third view, which Stanovich calls the Apologist, has it that although humans fall short of the norms of rationality, this is only because humans are not capable to meeting that standard due to their finite cognitive capacities. The "Apologist admits that performance is suboptimal" because humans are not perfectly normatively rational, but they are not *irrational* because there is no gap between how humans ought to act (given their limitations) and how they actually act (Stanovich 1999, p. 28). In his recent manuscript, *Rationality and the Reflective Mind* (2011), Stanovich drops talk of the Apologist and focuses on the Panglossian and Meliorist positions. By focusing on individual differences (i.e. the fact that not all subject respond in the same way to experiments) and advocating a tripartite division of the mind, he hopes to reconcile these two disparate views (2011, p. 10).

Stanovich (1999) surveys the existing two-system literature of the time, focusing on Sloman (1996), Evans (1996), Evans and Over (1996), Pollock (1991), and Levinson (1995). Because Stanovich's terminology has been so influential, it is worth quoting the passage in which he introduces the terminology at length:

> "Although the details and technical properties of these dual-process theories do not always match exactly, nevertheless there are clear family resemblances…In order to emphasize the prototypical view that is adopted here, the two systems have simply been generally labeled System 1 and

[1] Davidson and Dennett have distinct arguments for the claim that humans are basically rational. The upshot of their argument for debates concerning rationality is that the question of whether humans are rational or not would not be an empirical question. Stanovich attempts to refute their arguments in his (1999).

> System 2. The key differences in the properties of the two systems are listed next. System 1 is viewed as encompassing primarily the processes of interactional intelligence. It is automatic, largely unconscious, and relatively undemanding of computational capacity. Thus, it conjoins properties of automaticity and heuristic processing as these constructs have been variously discussed in the literature. System 2 conjoins the various characteristics that have been viewed as typifying controlled processing. System 2 encompasses the processes of analytic intelligence that have traditionally been studied in psychometric work and that have been examined by information-processing theorists trying to uncover the computational components underlying psychometric intelligence." (Stanovich 1999, p. 144).

The idea of a family resemblance is one that Stanovich has continued to claim. For example, he says: "My purpose here is not to adjudicate the differences among the [two-system] models. Instead, I will gloss over differences and instead start with a model that emphasizes the family resemblances" (2011, p. 17, see also p. 92). Stanovich includes a table similar to mine below. The properties he lists for S1 are as follows: associative, holistic, automatic, relatively undemanding of cognitive capacity, relatively fast, acquisition by biology, acquisition by exposure, acquisition by personal experience, highly contextualized, personalized, conversational, socialized, and interactional (i.e. conversational implicature). In contrast, he lists the following for S2: rule-based, analytic, controlled, demanding of cognitive capacity, relatively slow, acquired by culture, acquired by formal tuition, decontextualized, depersonalized, asocial, and analytic (i.e. psychometric IQ). Stanovich claims that differences in cognitive ability will only be found in problems that cue both S1 and S2. In (2011) he goes on to claim that differences in cognitive ability (specifically thinking dispositions) are differences in a certain kind of Type-2 process (more on this in a moment). Stanovich claims that S1 is associated with what Levinson (1995) called interactional intelligence (i.e. evolutionary intelligence), while S2 is associated with SAT scores and is identified with psychometric g. Most importantly, Stanovich says, the triggers for S1 are "highly contextualized, personalized, and socialized," while S2's "more controlled processes serve to decontextualize and depersonalize problems" (1999, p. 146).

Stanovich emphasizes two ideas in his work on rationality: individual difference and, what he calls, the 'fundamental computational bias.' The fundamental computational bias occurs when thinkers frame a problem within a familiar context. It is primarily carried out by S1, and earns its name from it pervasiveness (1999, p.192). Stanovich explains that the fundamental computational bias is "no doubt rational in the evolutionary sense" (1999, p. 202), though not the normative sense. In his 1999 monograph he uses this bias to argue that there are reliable

deviations from the norms of rationality, and so the Panglossian is wrong (1999, p. 208-9). He claims that we have greater predictive power if we relax "the idealized rationality assumption of Dennett's (1987, 1988) intentional stance by positing measurable and systematic variation in intentional-level psychologies" (1999, p. 212). Stanovich reiterates these points elsewhere (2011, p. 44).

Table 1.1: The Standard Menu (Most typical properties in bold)

System 1	**System 2**
Fast/ Parallel	**Slow**/ Sequential
Heuristically (or Associatively)-Based	**Rule-Based**
Non-linguistic	Linguistic (or sometimes dependent upon language)
Unconscious/ preconscious	Conscious
Implicit/ Tacit	**Explicit**
Mostly Shared with non-Human Animals	Uniquely Human
Evolutionarily Old	**Evolutionarily New**
Modular	**Unified**
Evolutionarily Rational	Individually Rational
Domain Specific	Domain General
Pragmatic/ Concrete Reasoning	Abstract Reasoning
Automatic	**Controlled, Volitional**
Subpersonal	Personal
Independent of Cognitive Capacities	Dependent upon Cognitive Capacities
Not Easily Altered	Malleable
Universal Among Humans	Varies by Individual and Culture
Independent of Normative Beliefs	Influenced by Normative Beliefs
A Set of Systems	A Single System
High Capacity	Low Capacity
Holistic/ Perceptual	Analytic Reflective
Independent of General Intelligence	Linked to General Intelligence
Independent of Working Memory	Liked to Working Memory
Stereotypical	Egalitarian

While many subjects respond incorrectly to tests from the heuristics and biases literature, they do not all get the answer wrong in the same way, and some get the answer right. This is what Stanovich calls 'individual difference.' Psychologists must explain this variation. Stanovich claims that the variation in thinking styles has not received the attention it deserves because

philosophers have focused on the competence/performance distinction such that "all the important psychological mechanisms are allocated to the competence side of the dichotomy" (1999, p. 213-14). Stanovich (2011) goes on to outline a taxonomy of reasoning mistakes and outlines why some, but not all, participants respond incorrectly. I will take up this issue in chapter 5.

Later, Stanovich moved (somewhat) away from defining S1 and S2 in terms of family resemblance. In fact, he moved away from talk of S1 and S2 altogether. This is because S1, on his view, is not a single system at all. Rather, it is a collection of module-like systems. While Stanovich (2009, 2011) is explicit on this point (calling the collection of module-like systems TASS for 'The Autonomous Set of Systems'), this view is implicit in his (1999): "Throughout the discussion of evolutionary rationality…it has been assumed that the computation underlying specifically evolutionarily adapted responses (as opposed to general problem-solving abilities) is accomplished entirely by System 1 modules and does not implicate analytic Intelligence." (1999, p. 238). He then cites Pylyshyn (1984) saying that these modules are hardwired and cognitively impenetrable, and that they evolved to deal with the environment. Members of TASS are like Fodor's modules in all ways except that they are not domain-specific.[2] Because S1 is, technically, a collection of systems, Stanovich moves to talk of Type-1 and Type-2 processes.

Stanovich (2009, 2011) also shifts the way he distinguishes the two kinds of processes. "The defining feature of Type 1 processing is its autonomy—the execution of Type 1 processes is mandatory when their triggering stimuli are encountered, and they do not depend on input from high-level control systems" (2011, p.19). Of course, Stanovich claims, a number of properties will closely correlate with autonomous[3] processes: they will be fast, will not use much executive functioning or central processing, and can operate in parallel. However, these properties are not essential to a process's being Type-1. Stanovich goes on to claim that autonomous processes include: "behavioral regulation by the emotions; the encapsulated modules for solving specific adaptive problems that have been posited by evolutionary psychologists; processes of implicit learning; and the automatic firing of overlearned

[2] Fodor himself might claim that this is an impossibility, since a number of the modules' properties are had in virtue of their being domain-specific, but we need not concern ourselves with that now.

[3] I take Stanovich and Evans' word choice of 'autonomous' to indicate members of TASS do not require the use of working memory, and so are 'autonomous' from working memory.

associations" (2011, p. 19-20). This way of identifying Type-1 processes begins to look like Stanovich's (1999) cluster proposal, since Type-1 processes are automatic, modular (and so evolutionarily old and fast), and heuristic. Stanovich defines Type-2 processing using the contrary of each property he used to define Type-1 processing. Thus, Type-2 processing is nonautonomous, slow, does put pressure on central computing, is serial (i.e. not parallel), and is often language based (2011, p. 20). All hypothetical thinking is Type-2 processing, though the inverse does not hold (2011, p. 47).

Perhaps the most important task of Type-2 processing is its capacity to override Type-1 responses. Stanovich claims that some of the inhibitory mechanisms carried out by Type-2 processing are "of the type that have been the focus of work on executive functioning" (2011, p.21), but not all override functions are of this type. Furthermore, while Stanovich claims that not all S1 overrides are efficacious (1999, p. 243), in his (1999) he talks as if they are. When does Type-2 processing fail to override Type-1 processing? And what other kinds of overrides are there such that they do not involve executive functioning? To answer these questions, we need a complete picture of the relation between the systems that carry out Type-1 and 2 processing. Stanovich (2011) offers just such an account. I begin by distinguishing Stanovich's systems and then turn to how they relate.

Stanovich (2009, 2011) moves from a two-system theory to a tripartite division of rationality; humans have three minds, rather than just one or two. Type-1 processing is carried out exclusively by TASS. Type-2 processing is carried out by two systems: the Algorithmic mind and the reflective mind. Let us examine each of these in turn.

Stanovich calls TASS the 'autonomous mind' because each system which is a member of TASS operates in a mandatory way. Deficiencies of TASS "often reflect damage to cognitive modules that result in very discontinuous cognitive dysfunction such as autism or the agnosias and alexias" (2011, p. 37). The processes carried out by the autonomous mind are "composed of affective responses; previously learned responses that have been practiced to automaticity; conditioned responses; and adaptive modules that have been shaped by our evolutionary history" (2011, p. 63). Note the similarity to the above list of Type-1 processes. Humans default to this kind of processing whenever possible, which Stanovich and others call 'cognitive miserliness.'

TASS is evolutionarily rational and evolved for the purpose of passing on genes. He cites Dawkins (1982) throughout his work.

The Algorithmic and reflective minds both carry out Type-2 processing. Thus, they will share a number of properties and so cannot be as neatly distinguished from each other as they are from the autonomous mind (2011, p. 34). The algorithmic mind is measure by fluid intelligence (IQ), while the reflective mind is measured by critical thinking tests (such as syllogistic reasoning). Tests that measure executive functioning tap the algorithmic mind rather than the reflective mind. Stanovich points out that tasks used to measure executive functioning (e.g. the Stroop Test) do "not require reflective control" in Stanovich's sense (2011, p. 57). As such, Stanovich prefers to say that tasks like the Stroop Test are a "measure of supervisory processes" (2011, p. 59). In contrast, the more a test measures how well a subject conforms to the norms of rationality, the more it measures the reflective mind (2011, p.42).

While the algorithmic mind does involve Type-2 processing, it is also prone to the fundamental computational bias. This is because the algorithmic mind defaults to using information that is readily at hand (more on this in a moment). Thus, the reflective mind is needed to overcome most biases. The reflective mind further sets goals and epistemically regulates the organism's beliefs. While breakdowns of the algorithmic mind include "general impairments in intellectual ability of the type that cause mental retardation" (2011, p. 37), breakdowns of the reflective mind will be at the personal level. Since it is a function of the reflective mind to ensure one's beliefs are "well-calibrated," and the reflective mind sets personal goals, the reflective mind is necessary for full rationality (2011, p. 38). The reflective minds of individuals differ in cognitive style: perhaps most importantly, how willing that subject is to form or alter his or her beliefs (2011, p. 35). Stanovich emphasizes that the reflective mind is defined in terms of *levels of control* rather than processing time or consciousness. "The term reflective mind is defined in terms of cognition involving high-level control change, not in terms of dictionary definitions of the term reflective (thoughtfulness, contemplation, etc.)" (2011, p. 79). He then admits that perhaps 'reflective' mind is a misnomer.

Stanovich provides the following structure of the tri-process model. I will make reference to his diagram of this structure (copied below) in my explanation. As previously mentioned, humans default to Type-1 processing. Stanovich (2011) emphasizes that the systems generally

act in concert, but that when differences do arise, the person will generally be better off using Type-2 processing, especially the reflective mind (see also 1999, p. 192). The function of overriding Type-1 processing is a feature of the algorithmic mind (arrow A). The reflective mind initiates this override (arrow B) by calling the algorithmic mind to take the autonomous mind offline. Stanovich insists that this is a two-step process, and supports his claim through individual differences: one can fail to recognize that Type-1 processing must be taken offline, which will result in one kind of heuristic response, or one can recognize that Type-1 processing must be taken offline yet fail to do so, which will result in a different heuristic response.

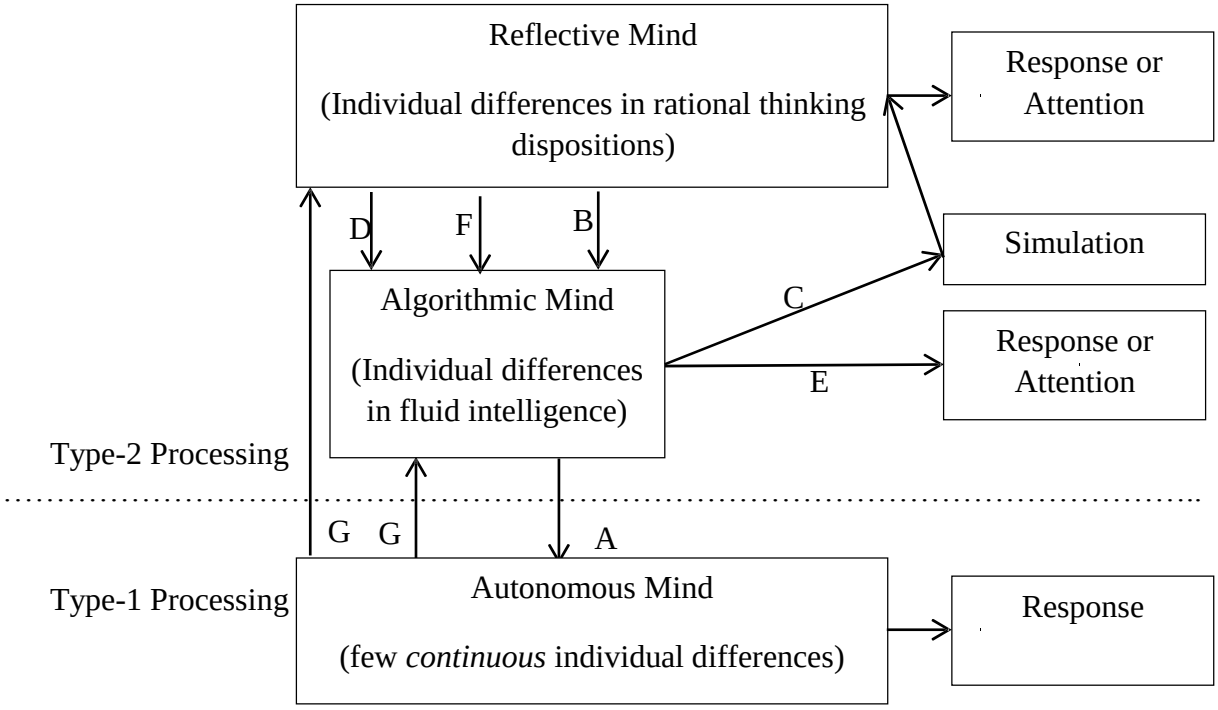
A simulation process is responsible for computing the alternative Type-2 response. The algorithmic mind carries out the decoupling (arrow C)—that is, the algorithmic mind abstracts from the details of the problem by forming a second-order representation of the problem (I will deal with this in more detail in chapter 5). The reflective mind issues the "call to initiate simulation" (p. 61) (arrow D). The reflective mind then uses the copy of the first-order representation to generate a response. Stanovich claims that the anterior cingulate cortex (ACC) "registers conflict during information processing and triggers strategic adjustments in cognitive control" (2011, p. 75). That is, it initiates the call for cognitive decoupling, and is therefore part of the reflective mind. It is the dorsolateral prefrontal cortex (DLPFC) that carries out sustained cognitive decoupling itself.

Strong decoupling is not necessary for Type-2 processing. Type-2 processing need not be hypothetical. The algorithmic mind can engage in serial associative cognition, resulting in either a response, or the organism's attention shifting (arrow E). When Type-2 processing is necessary, the organism defaults to "serial associative cognition with a focal bias (*not* fully decoupled cognitive simulation)" (2011, p. 66), which is "analytic (as opposed to holistic) in style, but it relies on a single focal model that triggers all subsequent thought" (2011, p. 67). Serial associative cognition tends to represent only one state of affairs (namely that which the subject takes to be the actual state of affairs), takes what is presented to the model as true, minimizes effort, ignores 'moderating factors' (e.g. possibility of social biases), and uses models the subject already believes or has used. The reasoning process of planning one's day is of this kind.

In addition to initiating an override of the Type-1 processing (arrow D), the reflective mind is also responsible for initiating an override of serial associative cognition (i.e. of the

algorithmic mind) (arrow F). It does so by preventing the algorithmic mind from continuing its process that would result in a new direction of attention or action. It may do so either by stopping the computation altogether in order to begin a simulation (arrow C), or it may begin a serial associative chain from a new starting point by "altering the temporary focal model that is the source of a current associative chain" (arrow E) (2011, p. 71).

Figure 1.1: The Tripartite Structure and the Locus of Individual Difference (from Stanovich 2011, p. 33 and p. 62)



1.2. Sloman

Steven Sloman claims that the two-system theory, at least in its functional (as opposed to anatomical) characterization has "gained widespread if not universal acceptance" (2014, p. 77). Sloman conceives of the two-system theory as a way of bringing together computational models and connectionist or associative models. In this way, two-system theories get the best of both theories: "only von Neumann components seem capable of manipulating variables in a way that matches human competences…yet associative components seem better able to capture the context-specificity of human judgment and performance as well as people's ability to deal with

and integrate many pieces of information simultaneously" (1999, p. 126, see also 1996, p. 3). Sloman (1996, 2002) claims that there is an associative system and a rule-based system, and he takes associative systems to be connectionist and rule-based systems to be manipulation of symbols, as in a language of thought (Fodor and Pylyshyn 1988). The "associative system encodes and processes statistical regularities" of the world (1996, p.3). These statistical regularities "approximate those of a sophisticated statistician" (1996, p. 4). The associative system computes based on "similarity and contiguity" (1996, 4) and the inferences it draws are "reflexive" (1996, 17). The associative system is not reproductive, in that it cannot freely recombine representations. However, it "can deal with novel stimuli" (1996, p. 16). The intuitive system operates on "reduced representations," by which Sloman means co-occurrence information, such as "statistical structure," rather than "logical structure" (2014, p. 72). The intuitive system is associated with emotions that are "directly tied to the perceptions of objects and events," such as anxiety, fear, and disgust (Darlow and Sloman 2010, p.382, see Sloman 2014, p. 72).

Sloman is one of the few theorists, and perhaps the only, who continues to maintain a token S1 position: S1 is not a collection of module-like systems. In defense of this position, he offers the following necessary condition on being a system: "a set of cognitive processes and representations must have some individual autonomy; they must operate and compute independently enough that they can be held responsible for critical aspects of behavior" (2014, p. 71). Sloman rightly says that we must constrain our concept of a system to maintain its explanatory power (2014, p.75). The difficulty, Sloman goes on to say, is that those who abandon the token S1 claim (such as Evans and Stanovich) fail to constrain their concept of a reasoning system.

While early on, Sloman (1996) claimed that the associative system does not operate based on causal or mechanistic structure, he later (2014) claimed that this was a mistake. Even in 1996, Sloman recognized that there was some relation between associative reasoning and causal structure. He claimed that associative reasoning was "sensitive to hierarchical and causal structure" because the associative reasoning depends "on representations constructed by rules, and those rules could construct different representations depending on hierarchical and causal knowledge" (1996, p. 16). However, this is not a very deep dependence. All it really amounts to

is the claim that observations depend upon the mechanics of the world. In 2014, Sloman altered his view, claiming that the intuitive system "represents the causal structure that produces statistical facts rather than representing the statistical facts themselves as associations. I also believe it cues us to action through tight links to affect" (2014, p. 77). Why did he change his mind? Consider the following two sentences:

Steven admires Daniel because he is so candid.

Steven annoys Daniel because he is so candid.

Notice that they are syntactically identical and differ in only one word. However, any native English speaker will immediately recognize that the 'he' of the first sentence refers to Daniel while the 'he' in the second refers to Steven. Sloman claims that this is because "'he' refers to the causal agent and that the sentences differ in whether the causal agent is the subject of the main clause (Steven) or the object (Daniel)…This suggests that causal structure is buried deep in our processing systems" (p. 72).

According to Sloman (1996), the deliberative system is rule-based. This second system is capable of representing the causal-mechanical structure of the world and utilizing these representations in reasoning (1996, p. 6). While the associative system can only represent finitely, the rule-based system is productive—it can represent an unbounded number of propositions. Sloman characterizes 'rule-based' broadly, to include formal rule-following accounts and mental modeling accounts of reasoning. The deliberative system requires the use of working memory (Darlow and Sloman 2010, p. 382, Sloman 1996, p. 17). The inferences it draws are "deliberative," which suggests that the deliberative system is a "goal oriented" system (1996, 17). However, Sloman later admitted that both systems are goal oriented (2014). The deliberative system is associated with emotions "that arise when alternative possibilities are considered," such as regret and frustration (Darlow and Sloman 2010, p. 382, see Sloman 2014, p. 72).

One major difference between the two systems is that of cognitive impenetrability. Subjects are always aware of the end result of a reasoning process, but they do not have conscious access to the process when it is solely an intuitive process (1996, p. 6, 2014, p. 77). However, he further claims that inaccessibility "provides only a fallible heuristic for identifying

systems, not necessary or sufficient conditions" (p. 6). He offers two reasons for this. First, both systems might be involved in forming a single response. Second, some rule-based processes might be unconscious. For example, mathematicians claim that the correct answer to difficult math problems leap to mind "even though their thoughts were elsewhere" (1996, p. 6). I should note that these considerations only refute the claim that conscious access to a reasoning process is *sufficient* for that process being associative (as opposed to *necessary*). Indeed, given his characterization of associations, it is unclear how an associative process could be consciously accessible. After all, in associative reasoning "concepts are represented by a set of features, so any features that the context brings along are automatically included in the concept representation" (1996, p. 8).

Table 1.2: Sloman's Standard Menu

Intuitive System	**Deliberative System**
Product is conscious, process is not	Agent is aware of both product and process
Automatic	Effortful and volitional
Driven by similarity and association	Driven by more structured, relational knowledge
Fast and parallel	Slower and sequential
Unrelated to general intelligence and working memory capacity	Related to general intelligence and working memory capacity

Sloman takes care to say how the two reasoning systems should not be characterized. First, the systems are not distinguishable by their domains. This can make it difficult to assess which system delivered a specific response. Second, the intuitive/ deliberative distinction is not the same as the distinction between induction and deduction (1996, p. 17). He explains that both can engage in inductive arguments (such as the inclusion fallacy) and deductive arguments (such as "belief-bias effects that are assessed, and in contradictory ways, by the two reasoning systems" (1996, p. 18)). Third, the distinction is not between conscious and unconscious processes, since the products (though not the processes) are available to consciousness in both kinds of reasoning (2014, p.70). However, I should note that intuitive reasoning processes are not accessible via consciousness while deliberative processes can be. Fourth, the distinction is not between rational and irrational processes. Sloman claims that deliberative processes can lead to

irrational conclusions when the wrong rule is employed,[4] and intuitive thought can be rational, as when identifying what is and is not food at the dinner table (2014, p. 70). Finally, the distinction Sloman has in mind is not the same as an analytic/nonanalytic cognition distinction (e.g. Allen and Brooks 1991), where analytic processing is a process that stores information abstractly. Sloman (2014) gives a table of properties of each system which he claims are common to all two-system theories (see above).

Sloman cites Lieberman saying that the intuitive system includes the lateral temporal cortex, the ventromedial prefrontal cortex, the dorsal anterior cingulate cortex, amygdala, and the basal ganglia. The deliberative system also includes the basal ganglia, but also includes the prefrontal cortex, the anterior cingular cortex, and the surrounding medial temporal lobe region (especially hippocampus) (see Lieberman 2007, Sloman 2014, p. 75 and Darlow and Sloman 2010, p. 382). However, these anatomical characterization of the two-system theory plays no role in Sloman's account of human reasoning, and he is willing to reject the claim that the two systems possess neurological correlates (Sloman 2014, p. 77). Thus, the two systems should be distinguished functionally rather than neurologically.

Sloman argues that the two systems operate in parallel. The two are in direct competition with one another, as opposed to the intuitive system shutting down when the deliberative system comes online. Thus, Sloman's view is a parallel-competitive model as opposed to a default-interventionist model (see Evans and Stanovich). Furthermore, Sloman claims that one can feel the tension between the two systems when they deliver contradictory outputs. He explains that "associative thought *feels* like it arises from a different cognitive mechanism than does deliberate, analytical reasoning" (1996, p. 3). Sloman takes this tension to be the best evidence for the two-system theory. He offers what he calls 'Criterion S,' according to which if a subject delivers simultaneous contradictory responses, there must be two systems at work. He offers the Müller-Lyer illusion as an example. Even after one measures the two line segments and finds them equal in length, they still *appear* different in length. This is good evidence that perception and reasoning are governed by two distinct systems (2002). Sloman claims that a similar case can be made for instances of reasoning. I will take up his argument in detail in chapter 3.

[4] It is unclear to me whether this would constitute an *irrational* response so much as an incorrect response.

Sloman is very clear about the relation that the two systems bear to one another: S2 suppresses S1, as opposed to there being some other system moderating the two (as in Evans 2009, Stanovich 2011, and Frederick 2005). Of course, S1 always has "its voice heard," since it is automatic (1996, p. 15). Sloman characterizes the functional relation between these systems saying that the "autonomous [intuitive/ S1] system operates through positive feedback between intuitive and affective components to relate the body to a pattern recognition processes. A second deliberative system operates independently and in parallel [to the first system], and serves to modulate the intuitive-affective loop via inhibition" (2014, p. 75). Sloman proposes that one loop of the frontal corticobasal ganglia loop, which acts as a gating mechanism, is the intuitive loop. Deliberation is an anterior prefrontal corticobasal ganglia loop, one function of which is "to serve to gate or at least modulate the intuitive-affective loop" (2014, p. 74). This way of characterizing the two systems does not, it seems to me, add anything to the functional description of their relation, and such an anatomical characterization is somewhat odd given that Sloman is willing to give up claims about the location of the intuitive and deliberative systems (2014, p. 77).

Sloman's views on the relation between S1 and S2 were, at one time, somewhat naïve. Even though he was always clear that both systems can affect each other, in his early work (1996) he seemed to assume that when the deliberative system conflicts with the intuitive system the deliberative system will succeed in supressing the intuitive response. This is not always the case. Sometimes the deliberative system will fail to suppress the intuitive system (Sloman, 2014). When will a deliberative response be overridden by an intuitive response? If deliberation does not lead to certainty, subjects will often rely on intuition, since it *feels* more right (2014, p. 74). The Cognitive Reflection Task devised by Frederick (2005) measures not only the ability for the deliberative system to come online, but also its ability to suppress the intuitive response. Note also that, while Sloman focuses on cases where the two reasoning systems offer different responses (since therein lies the best evidence for the two-system theory), the two systems can cooperate on tasks. Indeed, they usually do.

Sloman does not use the common S1/S2 terminology. In 1996, the terminology had not yet been introduced. However, in 2014 Sloman argues that we should abandon the S1/S2 distinction. He says that this distinction is "misleading if taken as a generic description of all

dual-systems theories in that they imply a single distinction when, in fact, different theorists have made different (if related) distinctions (as noted by, for instance, Evans, 2009, and Stanovich, 2011)" (2014, p. 70). He opts instead for an intuitive/ deliberative distinction. However, we must ask how we define these intuitive and deliberative processes. As it turns out, Sloman continues relying on the rule-based/associative distinction. For example, he writes "rule-based thought can lead to rational conclusions…but it can also lead to irrational conclusions when the wrong set of rules is systematically applied" (2014, p. 70). Perhaps what Sloman and others mean, when saying that the distinctions dual-system theorists have in mind differ, is that the dual-system theories arising from various domains of psychology differ (e.g. learning: Lieberman 2007 and Reber 1993, reasoning: Evans, Stanovich, Sloman, Kahneman, and mind-reading: Apperly 2010 and Apperly and Butterfill 2009), rather than the claim that dual-system theories of reasoning have different distinctions in mind.

In summary, early on, Sloman maintained an associative/rule-based distinction to distinguish the systems. While Sloman downplays this distinction in his subsequent work, it is still in his theory. According to Sloman, there are exactly two token and type reasoning systems, which operate in parallel and compete with one another. However, while the intuitive system operates whenever the subject is conscious, the deliberative system need not always come online.

1.3. Evans

Jonathan St. B. T. Evans has offered much data in support of the two-system theory. Since I will turn to his arguments for two-system theory in chapter 2, my purpose here is not to recount his work on belief bias and human reasoning and how he thinks it supports the two-system theory. My purpose here is only to recount the details of Evans's version of the two-system theory.

Although a two-system framework is evident in his work as early as 1984, it was not until 1996 with the publication of *Reasoning and Rationality* that Evans, with David Over, offered the details of a two-system framework. Evans and Over (1996) attempted to synthesize data from the deductive and probabilistic reasoning and decision making literature. In that volume, they were aware that the two-system theories being employed by psychologists were in danger of being in conflict with one another and they hoped to build a theory compatible with the two-system

theories of other psychologists. Evans and Over argue that what they call the 'personal system' (which corresponds to S1) is characterized as being goal directed, while what they call the 'impersonal system' (which corresponds to S2) engages in abstract thinking. Recall that Sloman (1996) suggested that S2 was goal directed whereas S1 was deliberative. So Evans and Over (1996) and Sloman (1996) are in opposition to one another on this point. Each system has its own rationality as well. Rationality$_1$ is acting "in a way that is generally reliable and efficient for achieving one's goals", while rationality$_2$ is acting "when one has a reason for what one does sanctioned by a normative theory" (1996, p. 8). The goal directed system (since that is its only end), uses heuristics. Indeed, rationality$_1$may have an agent assert contradictions or make illicit inferences, as long as doing so helps the agent achieve her goal. 'Rationality$_2$,' on the other hand, is governed by the rules of logic and other normative theories of decision-making (1996, p. 8). Evans himself later claimed that his and Over's attempt to synthesize the two-system theories of psychologists failed (2009), and there is no talk of rationality$_1$ and rationality$_2$ in any of his work subsequent to 1996.

Evans has offered two widely cited reviews of dual-process theories of reasoning (2003, 2008). In them, he claims that, while the idea that there are two kinds of reasoning is old, the two-system theory is a "striking and strong claim that there are two quite separate cognitive systems underlying thinking and reasoning with distinct evolutionary histories" (2003, p. 454). These dual-process theories "abound" in psychology (2008, p. 256, reiterated in 2009, p. 33), and are radical because "dual-process theories of thinking and reasoning quite literally propose the presence of two minds in one brain" (2003, p. 458, see also Evans 2010a). Evans explains that S1 is shared with non-human animals, is a set of subsystems with "some autonomy," includes instinctive behaviors which are innate (includes "any innate input modules of the kind proposed by Fodor"), is "most often described" as being associative (i.e. consisting in neural networks), is "rapid, parallel,… automatic[, and] only their final product is posted in consciousness" (2003, p. 454). S2, on the other hand, is evolutionarily new, slow, serial, uses central working memory, allows for abstract hypothetical thinking, is important for novel cases (since S1 only draws on past experience or innate operations), is volitional, and is responsive to verbal instructions (2003, p. 454-456). Evans (2008) reiterates the distinction with a limited list: "almost all authors agree

on a distinction between processes that are unconscious, rapid, automatic, and high capacity, and those that are conscious, slow, and deliberative" (2008, p. 256). He later adds that S1 processes are "concrete, contextualized, or domain-specific, whereas System 2 processes are abstract, decontextualized, or domain-general" (2008, p. 261).[5] Evans also provides a comprehensive table of properties used to distinguish S1 from S2 (or Type-1 from Type-2 processes).

Evans (2008) admits that some of the properties used to distinguish S1 from S2 should be abandoned. Indeed, it is "not possible" to link S1 and S2 to all the properties on Evans's elaborate Standard Menu (2008, p. 270). He is most critical of the division of S1 and S2 using evolutionary age, claiming that many theorists have merely asserted that S2 is evolutionarily new. Furthermore, it is problematic to hold that all animals share one form of implicit reasoning, because modules will vary from species to species (2008, p. 260). He suggests that evolutionary age needs to be revised or abandoned as a way of distinguishing the systems, and seems to adopt the former, replacing 'evolutionarily new' with 'uniquely developed in humans'. He is also critical of the heuristic/analytic divide, citing that it contrasts with parallel/interactive divide, though many features are the same, such as fast, automatic, belief-based, as opposed to slow, sequential, and effortful (2008, p. 263). Given the way that the properties on the Standard Menu combine, "System 2 appears to be a more coherent and consistent concept in the generic dual-system theory than does System 1 because multiple systems of implicit cognitive processes exist" (p. 263).[6]

Although Evans was an advocate of the two-system theory, he now claims that it is better to talk of Type-1 and Type-2 processing, "since all theorists seem to contrast fast, automatic or unconscious processes with those that are slow, effortful, and conscious" (2008, p. 270 see also 2009). Type-1 processes are automatic, in that they are mandatory. Type-2 processes, on the other hand, "require access to a single, capacity-limited central working memory resource…[which] implies that the core features of Type-2 processes are that they are slow,

---

[5] While one might worry about using such an elusive concept as consciousness to distinguish these kinds of processes, Evans defends its use by stating that "while the problem of what consciousness *is* may seem intractable, the study of its function and evolution seems more promising" (2008, p. 258). The operational definition of consciousness that "seems to have appeared" in the work of two-system theorists is that S2 "requires access to a central working memory system of limited capacity, whereas System 1 does not" (2008, p. 259). Furthermore, conscious processes are "inherently slow, sequential and capacity limited" (2008, p. 258 see also p. 261).

[6] Here 'dual-system theory' is intended to be broader than just reasoning systems, so as to include the theories in other areas of psychology such as social psychology.

sequential, and capacity limited" (2008, p. 270, see 2009, p. 38). However, Evans (2009) is quick to point out that being slow, sequential, and capacity limited are not necessary conditions on being a Type-2 process. Being automatic is necessary and sufficient for being a Type-1 process, while being non-automatic is necessary and sufficient for being a Type-2 process. There may be a clustering as a consequence of these features, but such a clustering is unimportant to the characterization of the types of processing. Interestingly, Evans does not altogether abandon the claim that humans have two reasoning systems. While S1 is a misnomer (since it is really a set of systems), he retains S2 and claims that it can operate with both Type-1 and Type-2 processes (2008, 2009).

Evans claims that even if theorists abandon the two-system theory, they can still maintain the two-mind theory. Rather than thinking of the unifying theory being the two-system theory, one should "describe this grand unifying form of dual-process theory as the 'two minds hypothesis'" (2009, p. 35). The evolutionarily new mind is that which is uniquely developed in humans. He claims that while humans and other animals both make decisions, "unlike other animals…we can base our decisions on simulation and imagination of their consequences" (2009, p. 39). The difference between the old and new mind is in degree rather than in kind. Evans (2009) defines 'mind' "as a high-level cognitive system capable of representing the external world and acting upon it in order to serve the goals of the organism" (2009, p. 35). However, in his popular work *Thinking Twice: Two minds in one brain* (2010a), he talks as though the two minds are different in kind. Indeed, given their characterization, it is difficult to see how they could be only different in degree.

Evans has drawn a helpful distinction between two kinds of dual-process theories. The first might be called 'default-interventionist' or 'sequential,' according to which, heuristic processes are pre-attentive and provide content to analytic processes rather than controlling behavior directly (2009, p. 45). Evans counts himself, Kahneman and Frederick (2002), and Stanovich among proponents of this kind of theory. He claims that, on the default-interventionist account, Type-1 processes solve the frame problem "by rapidly and effortlessly contextualizing our thoughts [and] retrieving stored memories and beliefs that are relevant to the current context" (2009, p. 45, see also 2008 p. 270-271). The second account, for which Sloman is his best example, has it that the two systems operate in parallel with one another, competing to control

behavior. Evans thinks that both are important, and that "parallel dual-process theories concern two systems that provide alternative routes to behavioral control, while sequential theories concern the interactions between working memory and its many support systems" (2009, p. 47). In his (2009) paper he attempts to reconcile the two rival theories to one another. One problem for the parallel conception concerns the way in which conflict cases are resolved, since S2 is slower than S1 yet is supposed to decide between the two responses. Evans suggests that there is a third type of process (Type-3) consisting in "resource allocation, conflict resolution, and ultimate control of behavior" (2009, p. 48). Type-3 processes are not to be confused with executive functioning or working memory. Rather, they regulate these things.

Evans is a default-interventionist dual-process theorist. While his early work posited two kinds of rationality, he has abandoned this attempt to reconcile various dual-process theories, opting instead for the more radical two-mind theory as a unifying theory of human rationality. While he has abandoned the S1-S2 terminology, his Type-1 and Type-2 processing are divided remarkably similarly to the Standard Menu, even though he claims that Type-1 processes are essentially automatic. Throughout this dissertation, I will focus on Evans's default-interventionist dual-system theory, but I will note where the shift to dual-process theory is important.

1.4. Frankish

I will begin by summarizing portions of Frankish's *Philosophy Compass* article on dual-process and dual-system theories of reasoning. Frankish explains that Type-1 processes are "fast, automatic, and non-conscious," often being described as "associative, heuristic, or intuitive" and Type-2 processes are "slow, controlled, and conscious," and frequently described as "rule-based, analytic or reflective" (2010, p. 914). In its two-system form, S1 and S2 are both multipurpose reasoning systems, S1 being evolutionarily old and S2 being evolutionarily new, peculiar to humans, and supporting abstract thinking that operates in keeping with our norms of rationality.

Although dual-process and dual-system theories in their modern form are only about 30 years old, they have historical roots. The philosophical distinction between intuition and reason gets at the "core of dual-process theory," and philosophers as far back as Locke used such a distinction (2010, p. 915). The rationalists' way of talking about animals evoked something like

S1 as well. Frankish cites Descartes and Leibniz in particular. There is a family resemblance between current dual-process and dual-system theories and theories of the unconscious among some continental philosophers, such as Schopenhauer, Nietzsche, and Freud. Frankish also cites Francis Galton as a pioneer of studies of the unconscious. He also claims that Freud's two systems are remarkably similar to contemporary theories. The unconscious is associative while the conscious is logical. Furthermore, the contents of the unconscious are inaccessible to the conscious mind, just as S1 processes are inaccessible to S2 (Frankish 2009, p. 92). However, contemporary theories part with Freud when he claims that the unconscious consists largely of repressed impulses and memories.

In their modern iteration, dual-process theories developed independently in four areas of psychology: learning, reasoning, decision-making, and social cognition. In the field of learning, Reber claimed that humans can learn both implicitly through association and explicitly through rules. Dual-process theories in reasoning came from discrepancies between subjects' performance and introspective reports on reasoning tasks. Further evidence came from a need to account for biased judgments of the validity in syllogistic reasoning. Dual-process and system theories came from theorists such as Smith, Collins, Chaiken, and Trope. Early work focused on persuasion, and theorists suggested that there was a default associative process contrasted with a cognitively taxing process that assessed the content received from another person. As in the reasoning literature, dual-process models were devised to account for discrepancies between "actual social behaviour and reported attitudes" (2010, p. 917). Psychologists studying race, such as Devine, argued that stereotypes are automatic, but that behavioral responses can be controlled by subsequent reflective processes. Dual-system theories emerged in the decision-making literature as well. Kahneman and Tversky's Linda problem is one of the most famous experiments that is routinely presented as evidence for dual-system theories. Kahneman and Tversky's early work implicitly contained a distinction between heuristic-based intuition and rule-based reasoning, which subsequently became explicit. On their account (similar to Evans's) humans default to S1, but S2 can intervene.

Frankish claims that there have been a number of dual-process theories in philosophy. He cites Dennett's distinction between belief and opinion, Cohen's distinction between belief and acceptance, and his own distinction between belief and superbelief (more on this in a moment).

Frankish claims that, among these three philosophers, what is common is that there are "two types of belief: one implicit, non-linguistic and associated with parallel, connectionist processing; the other explicit, language-involving, and associated with serial, rule-governed processing" (2010, p. 919). Frankish concludes that folk psychology might obscurely track dual-system theory. That is, folk psychology somewhat assumes dual-system theory (or something like it).

In 2006 Frankish and Evans organized a workshop on dual-process and system theories, which culminated in Evans and Frankish's (2009) edited volume on the subject. Frankish notes some important contributions from that volume in his (2010). First, the processes S2 carries out are not always abstract and can fail to meet the standards required by normative rationality. Some theorists (such as Stanovich and Frankish) have adopted accounts of S1 and S2 that rely on core properties, "demoting the others to the status of typical but non-essential ones" (2010, p. 921). Some (such as Evans) suggest moving away from talk of systems and toward talk of processes. Evans and Stanovich suggest that the mind might be divided into three parts instead of only two.

Frankish (2004) offers a somewhat different kind of dual-systems theory, though he claims that empirical data from psychology still supports his theory. Furthermore, he claims that his theory has strong ties to two-system theories in cognitive and social psychology. Frankish (2004) is concerned with the metaphysics of belief. The term 'belief' comes to us from our folk psychology and, though it always picks out doxastic states, these doxastic states can vary a great deal. Those who want to include the folk concept in our scientific psychology (such as Fodor), he calls integrationists. However, one might argue that belief, as it is currently used by the folk, should be discarded or heavily revised (Frankish cites Churchland and Stich as eliminativists, and Clark, Dennett, Horgan, and Graham as revisionists). There is a further divide between representationalists and interpretivists about belief.

To find a way between these camps, Frankish argues that our folk psychological term 'belief' refers to two distinct kinds of mental states, which he dubs 'belief' and 'superbelief'. 'Belief' is "non-conscious, partial, passive, and non-verbal" (2004, p. 4). 'Superbelief' is 'flat-out, active, and often language-involving" (2004, p. 5). He further suggests that something similar could be the case with other mental states and argues that there are two distinct minds—

what he calls ‘mind’ and ‘supermind’. Mind is characterized by what he calls the ‘austere’ account of mentality which characterizes folk psychology as “a shallow theory, which picks out behavioural dispositions and offers explanations that are causal only in a weak sense” (2004, p. 5). On this austere reading of folk psychology, the mental is constrained by rationality, specifically Bayesian decision-making: rationality is a constraint we impose upon the mind to make sense of behavior. Thus “actions are guaranteed to be rational at the basic level” (2004, p. 147).

In contrast, the supermind is characterized by a rich interpretation of folk psychology, where folk psychology is take as identifying “functional sub-states of the cognitive system” and offers robust causal explanations (2004, p. 5). The supermind is a “premising machine” realized by the mind. Similar to Cohen’s (1992) ‘acceptance,’ to superbelieve that p is to take as a premise that p (in at least some contexts), while to simply believe that p is to endorse that p in all contexts (2004, p. 130). For example, a depressed patient might accept (i.e. superbelieve) that she is physically strong (perhaps at the advice of her therapist to think positively). This might lead her to exercise more frequently and to get out of her depression. However, if she found herself in the unlikely position of having to jump a long distance from one cleft of a rock to another, she would evaluate whether or not she could make the jump based on her actual physical ability. The patient does not believe that she is physically strong; she superbelieves it. In order for her to believe it, she would need to adopt it in all contexts.

Frankish has continued to hold a dual-attitude account of the mind. Frankish (2009) reiterates the distinction between belief and superbelief (p. 103-105). Frankish (2012) defends his dual-attitude account against the claim that an action-based account of S2 implies that there are no S2 mental states (Carruthers 2011, chapter 4, (more on this below)). On an action-based account of S2, “S2 events influence action by way of higher-order S1 beliefs” (Frankish 2012, p. 45). Furthermore, when an agent forms a belief that one judges that p, that agent commits herself to acting as though p is true. Finally, decisions and judgments terminate reasoning processes—they are the end result. On this much Carruthers and Frankish are in agreement. Frankish goes on to say that if an agent interprets herself as believing that p, then that subject commits herself to reasoning as though p were true. Frankish relativizes the claim that decisions and judgment terminate reasoning processes—S1 (or S2) decisions and judgments terminate S1 reasoning

processes (or S2 reasoning processes, respectively). S2 attitudes are supposed to be realized in S1 attitudes. Suppose an action is mentally rehearsed and a higher-order S1 belief is formed about the action. Where then is the S2 belief—the superbelief? Frankish tells us that the S2 attitude would be "realized by a combination of the rehearsed utterance and the resulting higher-order S1 belief" (2012, p.47). Importantly for his defense against Carruthers, this resulting S2 belief need not be occurrent (p. 47).

Because there are two kinds of belief, and two kinds of mind, there will be two theories of mind: austere and rich. Frankish seems to think that this is a common-sense way of thinking about the mind. He repeats this common-sense justification elsewhere, citing conflicts between conscious and unconscious beliefs such as aversive racism (2012, p. 49).

An odd implication of this way of dividing things is that irrationality is located entirely at the S2 level. Because mind (which corresponds to S1) is to be analyzed in an interpretationalist way, it is guaranteed to be rational. Frankish (2004) endorses this claim: "given an austere view of the basic mind, our actions are *guaranteed* to be rational at the basic level, since *attributions of basic attitudes are constrained by an assumption of rationality*" (2004, p. 147, emphasis added). Thus, any instance of irrationality must be due to the supermind (which corresponds to S2). However, this way of dividing things is odd because (as we have seen above), S1 is supposed to be the heuristic, fast, intuitive, associative system that is less in keeping with our full normative principles of rationality. Compared to cognitive psychologists, Frankish's way of dividing rationality is reversed. This would not be so odd if it were not for Frankish's insistence that his work on the metaphysics of belief is supposed to fit nicely with (and is even supported by) two-system theories of reasoning.

Could it be that S1 corresponds to supermind and S2 corresponds to mind? No. Apart from Frankish's idiosyncratic way of dividing rationality between the systems, the functional properties of mind and supermind fit more naturally with S1 and S2 respectively. Superbelief is voluntary and belief is involuntary, while S2 is a controlled system and S1 is an automatic system(s). Mind and belief are non-conscious, which is how S1 is sometimes characterized, while supermind is conscious, which is how S2 is sometimes characterized. Thus, it would seem that we ought to understand mind as corresponding to S1 and supermind as corresponding to S2, even though Frankish's placement of irrationality differs from typical two-system accounts. Note

that automatic/controlled and non-conscious/conscious were two of the property pairs Frankish (2010) himself says are used to distinguish the two systems. Again, I must stress that Frankish emphasizes his continuity with two-system theorists.

Frankish's more recent work on two-system theory (2009, 2012) has been more in line with typical conceptions of S1 and S2, though some differences are still evident. Frankish reiterates that there is "abundant evidence" (2009, p. 89) and "converging lines of argument from different disciplines" (2012, p. 42) supporting the claim that there are two kinds of processing in "human reasoning, decision making, and social cognition." Thus, he takes his theory to be part of the same project upon which cognitive and social psychologists are working. Frankish (2009, 2012) agrees that S1 is a collection of systems. Frankish (2009) suggests that the personal/ sub-personal distinction can answer a number of pressing questions for two-system accounts. (For example, what is the relation between S1 and S2? Are there distinct memory systems as well? Do these systems share resources?) Briefly, Dennett's distinction between the personal and sub-personal has to do with attribution. A personal level state/process/event is one that is ascribable to the person or creature as a whole. A sub-personal level state/process/event is one that is not ascribed to the person or creature as a whole, but is rather ascribed to a part (or a subsystem) of that person or creature.[7]

Frankish suggests that we identify S1 with sub-personal level attribution and S2 with personal level attribution. He gives us the following examples for personal and sub-personal reasoning. Suppose you are asked what is 21,582 divided by 11. If you are a math whiz, the answer may just come to you (1962). You would not, however, know how you *worked out* the answer. The process of determining the answer would be entirely sub-personal. However, most of us would need to get out a pencil and paper and work through a series of steps. This process would be personal (though some steps along the way might be sub-personal (e.g. what is 22 divided by 11)).

The "defining feature" of personal reasoning is intentionality, by which Frankish merely means acting for reasons (2009, p. 92). Personal reasoning requires the use of working memory

[7] 'Person' should be understood in a very minimal sense. Personal-level states are not sufficient for personhood and do not themselves constitute the 'self.' Frankish is clear that he does not wish to imply otherwise (2009, p. 91).

and is "therefore conscious" (2009, p. 93).[8] However, the beliefs and desires motivating a particular instance of personal reasoning need not be conscious (i.e. they can be implicitly held). Personal reasoning is "slow, controlled, effortful,…conscious,…serial, shaped by culture and formal tuition,…demands attention, [and is] demanding of working memory" (2009, p. 96). Sub-personal reasoning is "typically fast, automatic, effortless, and non-conscious" (2009, p. 96).

Assuming that the sub-personal/personal distinction maps neatly onto the S1/S2 distinction, Frankish notes some important implications. First, S2 would not be a neural system in its own right, but is, rather, a virtual system "constituted by states and activities of the whole agent" (2009, p. 97). It is constructed out of sub-systems (2009, p. 99). He calls this an action-based view of S2 (2012, p. 42). Second, S2 is causally and instrumentally dependent on S1: instrumentally because S2 will use S1 subsystems to engage in autostimulation, whether it be inner speech, action simulation, or something else, and causally dependent because S1 (the sub-personal systems) generates the intentional actions used by personal reasoning. Lastly, S2 depends on S1 "to make its *outputs* effective" (2009, p. 97). That is, sub-personal "metacognitive attitudes make personal decision effective" (2009, p. 98).

Frankish's dependency claims have important implications for how S1 and S2 outputs override one another when they conflict. Suppose S1 generates a desire to perform action X, and that this desire enters into personal reasoning, resulting in a desire to perform an action incompatible with X (call it Y). Now, the subject has a desire to act in accordance with her personal desires. Thus, there is a conflict between a first-order desire to perform action X and a second-order desire concerning Y. Frankish says that whichever desire is stronger (the first-order or second-order desire) will win out, causing the subject to engage in the relevant behavior (2009, p. 99). Thus, Frankish has an elegant answer to how conflicting S1/S2 responses are resolved, even if desire strength stands in need for further analysis.

1.5. Carruthers

Peter Carruthers did not begin publishing on two-system theories until 2009, but a two-system account fits nicely with his previous work, especially his 2006 book *The Architecture of the Mind.* Since then, he has repeatedly claimed that the two-system theory is widely accepted

[8] Why working memory implies consciousness is unclear.

(see, for example 2013c, p. 339; 2013b, p. 236). For the most part, Carruthers divides S1 and S2 in a typical fashion, claiming that "System 1 is really a set of systems that are swift, implicit, unconscious, and largely shared with other animals, issuing in hard-to-eradicate intuitions about the answers to reasoning problems. System 2, in contrast, is uniquely human, and is slow, explicit, conscious, and controlled" (2013b, p. 236). Elsewhere Carruthers states that S1 is "heuristic in nature ('quick and dirty'), rather than deductively or inductively valid" and that S2 is serial (2009, p. 110). Carruthers cautions that theorists should be careful in what they classify as an S2 task. An S2 task requires either "recall and rehearsal of some appropriate culturally-acquired item of information (e.g. a normative belief),…the controlled activation of sequences of mental rehearsal in accordance with learned rules, or they should implicate practices of self-interrogation" (2009, p. 124).

In contrast to typical two-system accounts, but in keeping with Frankish, Carruthers offers an action-based account of S2. That is, S2 is realized in the cycles of S1. How is this supposed to work? Action-schemata are mentally rehearsed using the motor command system, but with the action schema taken offline (i.e. the set of instructions for muscle movement is suppressed) (2009, p. 113). In so being rehearsed, the representations involved are globally broadcast (and so made available to all the intuitive systems).[9] Having received new information, the intuitive systems then respond and output new cognitive and affective states that influence what will be rehearsed next. Inner speech is also important. A speech act is one kind of action. Thus, inner speech will still use the action-schemata. A speech action-schema is formed by the language module in light of one's beliefs and desires, but the speech act performance is supressed. Instead, a reference copy of the motor instruction, which would result in the speech act, generates an auditory representation that is then broadcast to the intuitive systems. Like overt speech, "we seem to hear the meaning of the imagined sounds of inner speech (the message expressed) as well as hearing those imagined sounds themselves" (2009, p. 117). (For a repetition of this account, see 2013a, p. 8).

Carruthers's motivation for positing this account is to answer a number of pressing questions for the two-system theorist: why would a second system evolve alongside the first?

[9] Note that only the contents of the representations will become conscious through the broadcasting, rather than the contents and the processes forming them (2009, p. 114).

How can there be a second class of mental states? Is there a second class of mental states? How do S1 and S2 relate and interact? Since actions result from reasoning, how can a reasoning system by constituted by actions?

S2 did not evolve *alongside* S1, but rather piggybacked on S2. The mechanisms necessary for S2 are evolutionarily ancient. All that had to evolve was a language system and "a disposition to engage in mental rehearsals of action on a routine basis" (2013a, p. 8). So the evolutionary question is answered. How are the two systems related? S2 is not a system with a neurological basis, but rather is a virtual system realized in the cycles of S1. Furthermore, S2 does not have its own distinct belief/desire/decision-making systems—it uses those mechanisms from S1.

Carruthers and Frankish have disagreed on whether S2 possesses distinctive mental states, and whether a virtual system can have its own mental states. This debate is central to whether S2 constitutes a mind. Carruthers claims that S2 "lacks the defining properties of a mind" because it "does not contain any beliefs, desires, or decisions" (2013b, p. 243). In other words, a necessary condition on being a mind is having distinctive mental states (beliefs and desires) that interact such that the interaction constitutes reasoning and decision-making (2013b, p. 242).What would it take for S2 to possess such propositional attitudes? For sensory-involving events in S2 to count as propositional attitudes, those sensory events must have causal powers like those of other propositional attitudes. Now, for the S2 event (which is identical to the putative S2 belief) to cause any action, requires that the meta-cognitive belief resulting from the S2 process be "combined with a source of meta-cognitive motivation" (2013b, p.242). S2 beliefs would have to interact with S2 desires. Likewise for decisions. Carruthers, citing Bratman (1987, 1999), claims that a decision is the kind of thing that settles what one does. But S2 would-be-decisions do not settle the event. In order to cause one to act, the processes carried out by S2 would need to be included in some further practical reasoning. Consider a putative S2 decision: after some S2 activity, I token 'I will go to the bank.' This content gets interpreted, and so broadcasted, but it will be the S1 beliefs that do all the causal work. Thus, there are not S2 attitudes or decisions.

In a recent article, Carruthers (2013a) has claimed that, while there is a real distinction between intuitive and reflective systems, there is not a real distinction between S1 and S2 (or

Type-1 and Type-2 processes). Carruthers assumes that there is a distinction between intuitive and reflective reasoning, and then argues that most of the properties on the Standard Menu cross-cut the intuitive/reflective distinction.

Carruthers argues that non-human animals engage in unreflective processes that can be flexible and rule-governed (2013a, p. 6). He does so, in part, by claiming that intuitive processes are not associative, and so must be rule based. In fact, he seems to argue for the stronger claim that no processes are associative. This should not be surprising, given Carruthers's commitment to computationalism. He cites Gallistel and Gibbon (2001) and Gallistel and King (2009) as demonstrating problems for associationists. Carruthers claims that, if associationism is true, then the number of reinforcement trials should be affected by whether or not they are interspersed with non-reinforced trials. However, the number of reinforcements necessary for animals to learn is not affected by mixing reinforced trials with unreinforced trials. So paradigmatic instances of associative processes are not associative. Thus, Carruthers concludes, intuitive systems must be computational (2013a, p. 5-6; see also 2009, p. 110). Furthermore, he concludes from this data on animal learning that the evolutionarily old/new distinction cannot be used to distinguish S1 from S2.

Next, Carruthers abandons two claims to which he had previously committed himself: that heuristics are 'quick and dirty' (2009, p. 110) and that S2's being action-based "explains why System 2 processes should be comparatively slow" (2009, p. 120). Carruthers says that heuristics can be 'rational' in the sense of ecological rationality (Gigerenzer 2000). That is, heuristics can be well-adapted in some environments but not well-adapted in others. In fact, intuitive reasoning can be better than reflective reasoning. For example, Wilson et al. (1993) offered subjects a choice between two posters (of the kind that college student might put in their dorm room). In one condition, they just chose, in the other they had to offer positive and negative properties of the posters before choosing. Those required to give positive and negative properties of the posters reported lower satisfaction with their choice a week later. Second, intuitive reasoning can be slow. For example, there are multiple cues in romantic relationships that are intuitive and slow: humans unconsciously detect the presence of pheromones and chemical information about their partner's immune system through saliva, and humans determine how kind their partner is (in part) given how they treat dogs and cats. A second example is that of

'sleeping on it,' where one puts a problem aside and the proper response later just comes to the individual. Carruthers claims that "it is natural to think that one must have continued reasoning about the problem, unconsciously, during the interim" (2013a, p. 11).[10]

Carruthers claims that intuitive reasoning is not always automatic because intuitive reasoning is goal-dependent (2013a, p. 13). Intuitive reasoning can be fully rational, since Gallistel et al. (2001) found that rats and pigeons could track randomly changing rates "as closely as it is theoretically possible to do" (2013a, p. 14). Carruthers also takes studies from Sperber and Mercier (2009) showing that people reason better in argumentative contexts to support the claim that intuitive reasoning can be normatively correct. This is because there is supposed to be an intuitive system designed for public argumentation.

Reflection can employ heuristics. For example, one can reflectively employ the 'sleep on it' heuristic, "which is often consciously and reflectively employed by people in our culture" (2013a, p. 16). However, this claim is odd, since earlier he called the 'sleep on it' heuristic "intuitive" (2013a, p. 11). Carruthers goes on to explain how many "culturally sanctioned reasoning heuristics" are not at all reliable (e.g. examining the entrails of a killed animal when making an important decision) (2013a, p. 17).[11] Finally, reflective reasoning is not unitary. That is, some are based on normative beliefs and lead to good reasoning when the beliefs are true, while others are skill-based and depend upon learned reasoning. Still others depend on modules that mediate transitions within cognitive architecture.

I must emphasize that Carruthers (2013 a) is not rejecting two-system theory; he is claiming that the two systems be divided in a way distinct from the Standard Menu. He tells us that "what really exists is a distinction between a set of intuitive systems and a reflective (mental rehearsal involving) architecture" (2013a, p. 19). Reflective reasoning uses mental rehearsal and global broadcasting, and will vary greatly by culture. He preservers some properties on the Standard Menu such as the unconscious/consciousness involving, being impervious to change versus malleable, impenetrable/penetrable by normative beliefs (2013, p. 19). In this dissertation,

[10] Because Carruthers (2009) committed himself to the view that S2's (or the reflective system's) being action-based explains why reflective reasoning is slow (as compared to intuitive reasoning), he now needs to show why an action-based account of the reflective system does not entail that reflective reasoning is slower than intuitive reasoning.
[11] It is unclear to me how this is a heuristic rather than merely reasoning with a false premise (if the entrails are such and such, then I ought to such and such).

I will engage mainly with Carruthers's views prior to the 2013a article, but I will address his novel way of distinguishing the two systems in chapter 4.

1.6. Summary

What do these theorists hold in common? First, they agree that humans possess more than one token reasoning system. Call this the **distinctness claim**. Second, these token reasoning systems are of two kinds. Call this the **kind claim**. Importantly, the kind claim implies that humans possess (at least) two token reasoning systems (i.e. the distinctness claim). Dual-system theorists go further and say how these two systems are individuated—they are individuated by a clustering of properties along the lines of the Standard Menu. In particular, they agree that S1 is fast, heuristic/associative, automatic, and evolutionarily old, and that S2 is slow, sequential, controlled, and evolutionarily new (or 'uniquely developed in humans' (Evans and Stanovich)). The account of kinds that these theorists seem to tacitly endorse is a homeostatic clustering account of kinds (see Boyd 1991, 1999). On this suggestion, S1 and S2, identified by a homeostatic clustering of properties, are supposed to constitute cognitive kinds. I will call the claim that the two systems are of different kinds classified by the Standard Menu, the **cluster kind claim**. Importantly, the cluster kind claim implies the kind claim, though the converse does not hold, since the kinds of reasoning system might be distinguished in some other way. Finally, most theorists claim that, while S1 is a collection of systems, S2 is a single system that operates across domains. Two-system theorists take it to be the case that it is the same (numerically) S2 operating in a variety of cases. Call the claim that S2 is a single, domain-general reasoning system the **token S2 claim**. I take these three claims to be the core of two-system theory. Some, such as Sloman, claim that there is one token S1 system as well (call this the **token S1 claim**). Although these three claims are closely related, and generally not distinguished explicitly in the literature, it is important to separate them, as some of the recent moves by Evans, Stanovich, and Carruthers involve altering some, but not all, of these claims.

Samuels (2009) distinguishes between two versions of two-system theories that many two-system advocates have adopted. The first is the token thesis, which states that each human has two *particular* cognitive reasoning systems. The second is the type thesis, which states that each human has two *kinds* of cognitive reasoning systems. Most two-system theorists implicitly

hold to the token thesis for S2 and the type thesis for S1. That is, two-system theorists think that S2 is a single system that is active across a wide range of domains, while S1 is a collection of systems. Again, Sloman is the exception.

Samuels claims that the token thesis implies the type thesis, but that the converse does not hold. Let us begin with Samuels's latter claim. If the type thesis is a thesis about types of *systems*, then indeed the type thesis would imply that there are at least two token systems (at least one of each type). However, the token thesis is the thesis that there is one system of that kind (i.e. there is just one S1 (which only Sloman holds) or there is just one S2 (which is part of the typical view)). Thus, the type thesis does not imply the token thesis about either system, since there may be many Type-1 systems and many Type-2 systems. The token S1 and S2 claims jointly imply that S1 and S2 are distinct types, if S1 and S2 are individuated (and so tokened) based on being distinct kinds. Confusingly, Samuels's distinction leaves out a good way to talk about the possibility that humans possess two reasoning systems of the same kind just, as humans possess two lungs of the same kind. Perhaps a better way to characterize the distinction is between having exactly two token systems (of two different kinds) and having more than two token systems of two kinds. This is the way I will use the type/token terminology.

Philosophers differ in how they distinguish the terms 'hypothesis' and 'theory.' Some think of a hypothesis as a local prediction—a prediction one makes for a single (or a small set) of experiments. Theories, on the other hand, unify many hypotheses, preferably confirmed hypotheses. On this view, hypotheses do not become theories over time, and there is a principled difference between the two. Another way to draw the distinction is to think of a hypothesis as mere educated guess, and a theory as a well confirmed hypothesis. This way of drawing the distinction says that hypotheses may become theories over time and might allow for border-line cases of claims that are somewhere between hypotheses and theories. I will use the former convention, according to which theories are that which generate hypotheses. However, I should note that this is not the way some of the two-system advocates I will be addressing in this dissertation use the terms. Two-system advocates freely move between talk of the 'two-system hypothesis' in one paper and 'two-system theory' in another. Evans and Stanovich (2013b) recently called dual-process theory a 'paradigm.'

Recently, some theorists have moved away from talk of two systems in favor of dual-processes (see especially Evans after 2008). Dual-possess theorists reject the token S1 and S2 claims, but retain a version of the kind claim (and perhaps the cluster kind claim), but with regard to processes rather than systems. Namely, they claim that there are two types of reasoning process divided along the lines of the Standard Menu, or some properties from the Standard Menu. It seems to me that this move does not alter the dialectic in a significant way. Systems are the kinds of things that carry out processes and are individuated *inter alia* (I will argue in chapter 2) by the kinds of processes they carry out. We should ask what kinds of systems carry out these two kinds of processes. It could turn out that dual-process theory collapses back into two-system theory. To avoid this, perhaps dual-process theorists wish to claim that there is just one reasoning system that operates in two kinds of modes, or that there are many reasoning systems, all of which can carry out the two reasoning processes. In this case, my objection to the kind claim will be all that is relevant to dual-process theory, but this is as it should be, since the kind claim is the only novel claim dual-process theory makes.

2. Philosophical reception of the S1/S2 distinction

Frankish (2010) rightly points out that "if our judgments and actions" are generated by one of two distinct mental systems, "then many traditional philosophical questions will need to be recast to allow for this duality, with implications for debates about agency, autonomy, responsibility, rationality and knowledge, among other topics," adding that this is "likely to be fertile area for future research" (Frankish 2010, p. 923). Indeed, a number of philosophers have adopted some form of the two-system theory and used it to their own philosophical ends.

Fiala, Arco and Nichols (2011) use the two-system theory to defend physicalism against the criticism that it cannot account for consciousness—that is, physicalism cannot bridge the 'explanatory gap.' They claim that ascriptions of a conscious state can be made in two ways. First, there is an intuitive 'low-road' process (originating in S1) that is sensitive to superficial features characteristic of agency. One example is greater willingness to attribute agency to inanimate objects when they have representations of eyes (see Johnson 2003). The 'low-road' accounts for our intuitions about consciousness. However, there is also a reflective 'high-road' process (originating in S2) that involves deliberate reasoning and draws on a wide variety of

information. Physcialist theories of the mind satisfy the 'high-road,' but not the 'low-road,' resulting in the illusion of a gap in explanation.

Jennifer Nagel adopts a dual-system account to defend single-premise closure, according to which, "if one knows P and competently deduces Q from P, thereby coming to believe Q, while retaining one's knowledge that P, one comes to know that Q" (Hawthorne, 2005 p. 29). However, suppose Muhammad Ali knows his car is parked in front of his house. Then, by single-premise closure, Muhammad Ali knows that his car has not been stolen. However, Muhammad Ali does not *know* that his car has not been stolen, since every day in Toronto some cars are stolen (see Vogel 1990). Nagel explains that "what is not immediately clear is why the deduction feels wrong" (2011, p. 2). She argues that bringing up alternative possibilities constitutes hypothetical thinking, which requires S2. However, when reasoning about routine cases (what Nagel calls 'lax propositions'), humans only use their S1 (because of cognitive miserliness). Since lax propositions are sufficient for knowledge, single-premise closure is preserved when using S1, though not necessarily when using S2.

Nowhere has the two-system theory been more widely applied than in moral theory. In particular, theorists have applied two-system theory either to find a way between sentimentalism and rationalism, or to defend rationalism. Jonathan Haidt's (2001) social intuitionist account of moral judgment claims that emotion plays the primary causal role in moral reasoning and so is an example of Humean sentimentalism. Rationalists, who think that deliberative reasoning is the cause of moral judgments, have replied that the sentimentalists have mischaracterized their position (see Kennett and Fine, 2008 and Smith, 2008). Consider, for example, Michael Smith (2004) who develops a rationalistic account of morality using moral concepts, especially the concept of reason for action. Assuming talk of reasons for action is right, humans do engage in systematic justification of their desires.

Joshua Greene et al. (2004) adopt a two-system account wherein there is a decision-making process that is evolutionarily old, emotional, and domain-specific and another system that is domain-general and abstract in its reasoning. S1 responses are supposed to ground the absolute prohibitions central to deontological accounts of morality, while S2 makes utilitarian reasoning possible (Greene et al. 2004, p. 398).

Jillian Craigie (2011) rightly points out that Greene's two-system account differs from Kahneman and Frederick's (2002). She claims that Kahneman and Frederick's account emphasizes the two systems working in an integrated way while Greene's account suggests that the two processes are fundamentally in competition with one another. That is, Greene's adopts a parallel-completive model rather than a default-interventionist model. Furthermore, Kahneman and Frederick allow that S1 processes can be altered, while Greene seems to suggest that they are fixed by evolution (though shaped and redefined by our culture). While Craigie is critical of Greene, she agrees with him on the need to incorporate the two-system theory into the moral judgment debate. Craigie also works with Kahneman and Frederick's (2002) account and says that although different two-system accounts have focused on "different distinguishing properties…essentially the distinction is one between cognitive operations that are fast and largely unconscious, and others that exhibit characteristics traditionally associated with controlled, reflective processing" (Kahenman and Frederick, p. 57). She concludes saying that both S1 and S2 are needed for competent moral decision making.

Ron Mallon and Shaun Nichols (2011) also use the two-system theory to argue against Haidt and defend moral intuitions by suggesting that S2 is responsible for moral justification and S1 is responsible merely for moral explanation. Take Haidt's famous incest example. Subjects were told that a brother and sister have consensual sex (taking contraceptive precautions) and the subjects were then asked whether the action was right or wrong. Subjects responded immediately that this was wrong, but when Haidt and colleagues challenged this claim subjects were very bad at justifying their answers.[12] Mallon and Nichols then claim that a subject might respond (using S1) that the brother and sister's actions were wrong. Mallon and Nichols claim that this is an explanation, but no justification whatsoever. Thus, what subjects are bad at is offering *justifications* for their moral beliefs.[13] This fits nicely with the human tendency of cognitive miserliness.

Two-system theories have already been applied in the fields of philosophy of mind, epistemology, and moral reasoning. If the more radical two-mind theory is right, this would have

[12] Though widely cited, Haidt's experiment has never been published. Problematically, Haidt required students to explain the harm in the siblings' action, which, arguably, assumes a consequentialist account of morality.
[13] Mallon and Nichols claim that both S1 and S2 can be rule using systems (which constitutes a change to the standard view).

profound consequences for our theories of responsibility and punishment, as actions can originate from either the old (i.e. S1) or new mind (i.e. S2). How can 'I' (which Evans 2010a says refers to the 'new' mind) be held accountable for what my 'old' mind did? There are important philosophical question upon which the two-system (and certainly the two-mind) hypothesis would have implications.

3. Cognitive kinds

If the two-system theory is to be understood as a theory of cognitive kinds, then we need to be clear on what a kind is. One plausible account of natural kinds (especially for cognitive science) is that of a homeostatic cluster of properties. Boyd (1999) takes biological species to be a paradigmatic case of natural kinds that are a clustering of properties. The clustering is due to a homeostatic mechanism. A homeostatic mechanism makes it the case that properties will remain in the cluster. A member of a species might be born with very divergent properties from the rest of their species, but (if these properties are too divergent) that member will be unable to reproduce and pass along its genes. Thus, when that member dies, its properties that vary from the typical cluster disappear, preserving the cluster of properties. In this case, part of the mechanism for keeping the properties in the cluster is the weeding out of members who bear divergent properties. However, the mechanism can also be such that it causes its members to bear certain properties.

If the kind claim is right, then S1 and S2 are kind terms. The various versions of the Standard Menu constitute different accounts of what properties should be included in the cluster of the kinds 'S1' and 'S2.' Thus, the homeostatic cluster account of natural kinds is well suited to evaluate the cluster kind claim. There is added importance for S1 with regard to the kind claim, since it is really a collection of systems. Within human cognitive architecture, we must have a way of delimiting the Type-1 systems from non-Type-1 systems.

4. The heuristics and biases themselves are the explanandum

Researchers in the heuristics and biases literature sometimes refer to biases as explanations for empirical data. One necessary condition on explanation is that it is an answer to

a why-question (van Fraassen 1980).[14] Why did the subject answer incorrectly to some problem, rather than getting it right? One might respond that this happened because the subject was using a specific heuristic (i.e. the representativeness heuristic, the anchoring bias, or the availability heuristic). Here I will argue that the identification and classification of various heuristics and biases do not explain (or are very minimal explanations of) the data from the heuristics and biases and reasoning literature. Instead, the individual heuristics, biases, and reasoning errors are themselves the explanandum while two-system and one-system theories are explanations. The relationship between an instance of a heuristic (such as Linda the bank-teller) to the general heuristic (such as the representativeness heuristic) is one of membership to class.

The goal of the heuristics and biases project, Kahneman and Tversky tell us, is "to understand the cognitive processes that produce both valid and invalid judgments" (1996, p. 582). So the heuristics and biases literature, in which we identify the ways in which subjects fail in reasoning, is supposed to illuminate the cognitive processes that lead to our judgments. However, simply identifying a specific heuristic does not accomplish this goal. To see why, we need to consider a specific heuristic as a putative explanation of an instance of that heuristic. Let us take Kahneman and Tversky's representativeness heuristic as an example. Briefly, the representativeness heuristic occurs when subjects respond in accordance with the answer that 'fits' with a description. For example, given a description of 'Linda' fitting well with her being a feminist, subjects are more likely to claim that she is a feminist bank-teller, rather than a bank-teller (Tversky and Kahneman 1983). Suppose we agree that people generally answer incorrectly in the Linda case because people are subject to a representative heuristic. Thus people generally answer incorrectly in the Linda case because, in cases like the Linda case, people respond according to whichever answer fits with the description. What does it mean to respond according to whichever answer fits with a description? It is to act in accordance with the representativeness heuristic. The explanation, then, is circular.

Cummins (2000) argues similarly, but more generally for all of psychology. He writes: "in psychology, such laws as there are almost always conceived of, and are even called, effects" (p. 199). However, "no one thinks the McGurk effect explains the data it subsumes. No one not

[14] van Fraassen claims that being an answer to a why question just is an explanation. However, only the weaker claim (that being an answer to a why question is necessary for being an explanation) is needed for my argument in this section.

in the grip of the DN [deductive-nomological] model would suppose that one could *explain* why someone hears a consonant like the speaking mouth appears to make by appeal to the McGurk effect. That just *is* the McGurk effect" (Cummins, p. 199). Here I am in total agreement with Cummins. What is odd is that a number of psychologists do seem to think that appeal to some heuristic explains an instance of that heuristic. Gigerenzer (2010) argues that this mistake is common in psychology, and that it is especially problematic within dual-process theories. He claims that, in dual-process explanations of belief-bias, "the phenomenon that both logical structure and prior belief influence judgments is 'explained' by restating the phenomenon in other words" (p. 738).

Perhaps saying that the subject responded incorrectly because they were using the representativeness heuristic offers some minimal explaining. Telling us that a heuristic is involved eliminates certain possibilities. For example, saying that the representativeness heuristic is involved tells us that the subject did not simply guess. However, this is a very minimal explanation, on par with the Moliere's doctor who tells us that opium causes one to sleep because of its soporific virtue. Perhaps this is no explanation at all, but, as David Lewis (1986) points out, it does rule out certain ways in which sleep could be brought about. For example, it rules out cases in which an agent secretly gives sleeping powder to everyone who takes opium. Saying that people respond incorrectly because they used the representativeness heuristic is similar in its explanatory strength. While saying that the representativeness heuristic is involved rules out possible explanations, those explanations are far-fetched and so do not provide a satisfactory answer to the question: why did the subject respond incorrectly?

While identifying the representativeness heuristic is not sufficient to explain individual occurrences of that heuristic, and so falls short of what Kahneman and Tversky had hoped to accomplish, it is an important step in developing an explanation. To see why, consider an analogy from biology. Suppose you ask why koalas carry their babies in pouches. Now suppose I tell you that this is because they are marsupials. If you have no idea what a marsupial is, it will be no good for me to point out that marsupials are mammals that have pouches. I may identify all the marsupials, and you may even come to learn what it is to be a marsupial. However, you still do not have a satisfactory explanation for why koalas carry their babies in pouches. A proper explanation will presumably have to do with evolutionary pressures on koalas (and marsupials

more generally) for the evolution of a pouch, or an account of their genetics. It is helpful to identify that koalas, in virtue of having a pouch (*inter alia*), are members of the marsupial infraclass, but it does no explanatory work (or at least very little). However, identifying 'marsupial' as a kind is important insofar as it gives us a taxonomy. The taxonomy clarifies the explanandum and elucidates what is relevant to koalas having a pouch. For example, their diet consisting of eucalyptus leaves is not relevant, since other members of the marsupial infraclass do not eat eucalyptus leaves.

Similarly to the relation between koalas and marsupials, the Linda case is an exemplar of the representativeness heuristic. Thus, saying that people use the representativeness heuristic in the Linda case is an unsatisfactory explanation. What we need to explain the Linda case (or at least begin to explain the Linda case) is a story about cognitive processes that underlie the representativeness heuristic. What we need is an account of the cognitive architecture of human reasoning. The two-system theory offers just such an account. I will argue that it is flawed and propose a one-system account to replace it.

5. Roadmap

In chapter 2, I will offer a brief sketch of a one-system alternative to the dual-system model. I offer some criteria for distinguishing systems to help clarify the dialectic between one-system and two-system theories. I will then argue that this one-system alternative can explain the data generally taken to support the two-system theory. In chapter 3, I argue that a one-system account of reasoning is incompatible with contradictory beliefs issuing from reasoning simultaneously. I argue, contra Sloman (1996, 2002), that we do not have sufficient evidence of such beliefs, but I propose experiments that might offer compelling evidence for their existence. Thus, a one-system account is empirically testable against parallel-competitive dual-system accounts. In chapter 4, I turn to the kind claim. After clarifying the properties most commonly found on the Standard Menu, I argue that these property pairs cross-cut one another. Thus, there is reason to think that the cluster kind claim is false. I then argue against monothetic ways of distinguishing the two kinds of systems from Carruthers, Stanovich, and Evans. Finally, I offer further detail of my own one-system alternative according to which there is only one flexible reasoning system that can operate in several of the ways outlined on the Standard Menu, but

which cannot operate in contrary modes simultaneously. I suggest a way to empirically distinguish my one-system account from default-interventionist dual-process theories (specifically those of Evans and Stanovich) and argue that there is evidence in favor of my account's prediction.

Chapter 2: Against the Inference to the Best Explanation for Distinctness

Advocates of the two-system theory must demonstrate that humans possess more than one token reasoning system (what I call the distinctness claim), that these reasoning systems are of different kinds (what I call the kind claim), that the two kinds of systems are distinguished by a clustering of properties along the lines of the Standard Menu (what I call the cluster kind claim), and that S2 is a single domain-general system (what I call the token S2 claim). Advocates of the two-system theory generally support the distinctness claim (or the distinctness claim combined with the kind claim) by an inference to the best explanation: there being two reasoning systems is the best way to explain experimental data from the reasoning and decision-making literature. In this chapter, I critically examine the inference to the best explanation for the distinctness claim in combination with the kind claim, concluding that an alternative theory can explain the data from the heuristics and biases literature just as well as the two-system theory.

Two-system theorists take themselves to be displacing a default cognitive architecture for human reasoning, yet they usually do not say what this existing alternative is, or how it would explain the existing data taken to support the two-system theory.[1] I begin by offering the outline of this alternative theory, which I call the one-system theory. Two-system theorists have also not always been clear as to what it is for two reasoning, cognitive, or mental systems to be distinct. I posit three principles for individuating systems. Although I have cognitive systems (or more specifically, reasoning systems) in mind, these principles generalize. Then, focusing on data from the heuristics and biases literature that has paradigmatically been taken to support the dual-system theory, I argue that a one-system account can explain the cases taken to support the dual-system theory. Thus, the inference to the best explanation for the dual-system theory is undercut.

[1] At times, two-system theorists do make gestures toward an alternative theory. Evans (2010a) offers the 'chief executive model' as an alternative theory. The chief executive model states that the conscious person is in charge of his or her decisions and reasoning processes. He points to Descartes as the progenitor of this view, but claims that it is a folk intuition as well. (He says the chief executive model is part of 'folk psychology'. By this, Evans means that the chief executive model is part of the folk's view of psychology, and he recognizes that he is using 'folk psychology' in a different way than philosophers and psychologists use it.) So the alternative to two-system (or in Evans's case, the two-mind) theory is folk intuition. We should consider a better alternative hypothesis.

1. An alternative theory: One-system reasoning

One possibility for the architecture of human reasoning is that there is only one token reasoning system. Call this the one-system theory. The one-system theory can and should allow that this system operates differently under different circumstances. The one system might have multiple modes of operation. For example, it can operate deductively and inductively. Note that this does not confuse the dialectic between the one and two-system theory. Two-system theorists emphasize that both reasoning systems engage in deductive and inductive reasoning (see Evans 2013a, Stanovich 2011). What the one-system theory denies is that there must be two token systems for there to be various kinds of reasoning. The various modes of operation, by which I mean 'ways of operating,'[2] would be properties of the single reasoning system—dispositions that manifest under specific stimulus conditions. The one reasoning system might operate consciously at times and unconsciously at other times. It might at times operate automatically and at other times operate under the conscious control of a subject (though this, I will suggest in chapter 4 and 5, is a matter of degree).

One may wonder if there is a substantive difference between the one and two system theories (see Keren and Schul (2009) for an explication of this worry). One might even claim that the division is merely a matter of semantics—a debate about how we should use the term 'system.' Perhaps the one and two-system advocates are speaking past each other, having different conceptions of 'system.' In order for there to be a substantive debate between the one and two-system theories, there must be an account of systems to which both parties can adhere. More specifically, there must be an account of system *individuation* to which both parties can agree. The principles of individuation, combined with the two-system (or one-system) theory, should generate empirical predictions such that these accounts are testable. If both parties can agree to some conception of 'system,' and one is a realist about reasoning systems, then there will be an ontological difference between the one and two-system theories. Namely, they will

[2] There is the possibility of confusion in my term 'mode.' I mean this term in the sense that analytic metaphysicians use it—modes are ways of being. Thus, modes of a process are ways that a process is. Recently, some dual-process theorists (Evans and Stanovich) have used the term to denote ways of operation within Type-2 reasoning. For example, on their use of the word, 'inductive' and 'deductive' are modes of Type-2 processing. They claim that modes are distinct from properties (see Evans and Stanovich 2013a). They seem to hold it in order to prevent Type-2 processing from breaking into further types. However, if properties are ways things (or processes) are, then 'modes' in Evans and Stanovich's sense do not turn out to be different than my sense.

answer the following question differently: are there two reasoning systems? Two-system theorists answer that there are two distinct entities of differing kinds, while one-system theorists claims only one. Two-system theorists use distinct entities to account for differences in reasoning; the one-system theorists use properties of a single entity to account for the same differences.

2. System individuation

Given that one and two-system theorists agreeing on how to distinguish systems is necessary for the debate to be substantive, it is surprising that two-system theorists have generally failed to be clear on how systems are to be distinguished. Keren and Schul (2009 p. 534) claim that two-system theorists generally rely on vague definitions. For example, Kahneman and Frederick (2002) say of S1 and S2 that "these terms may suggest the image of autonomous homunculi, but such a meaning is not intended. We use the terms *systems* as a label for collections of processes that are distinguished by their speed, their controllability, and the contents on which they operate" (p. 51). It will be helpful to have a more rigorous standard, and to that end I will propose three principles of system individuation to which both parties can agree. In chapter 3, I will examine how these principles affect empirical predictions.

I begin with some concrete examples. The digestive and the circulatory systems are distinct token and type systems, though they are causally related. Consider one example of their causal relation. If I consume whiskey, the alcohol is absorbed in the stomach and digestive track, entering into my blood stream and allowing my circulator system to distribute the alcohol throughout my body. Thus, the digestive and circulatory systems are not causally isolated from one another. Each system can be characterized functionally. The digestive system is the system that breaks down nutrients from food, and the circulatory system (*inter alia*) delivers oxygen and nutrients to muscles. There are functional and structural differences between these two systems, and either of these differences are sufficient for their being distinct token and type systems. Also note that systems can have subsystems. For example, the lymphatic system is a part of the circulatory system. The lymphatic system is part of the circulatory system mereologically (since all of the lymphatic system's physical parts are parts of the circulatory system) and functionally

(since the lymphatic system carries out a specific part of the function of the circulatory system). Finally, the lymphatic system makes up part of the circulatory system's structure.

The first thing to note from these examples is the importance of function in understanding systems. Indeed, I take systems to be things that carry out a function or functions. One might object to this characterization of systems citing systems that do not have any goal (such as the water cycle). However this objection assumes a certain view of functionality. Functions can be understood teleologically or causally/dispositionally (for the former see Millikan 1984, Neander 1991, and Wright 1976. For the latter see Shoemaker 1980, Bigelow and Pargetter 1987, and Martin 2008). The water cycle can be understood as functional in the causal/dispositional sense, which requires no 'goal'. Processes are closely related to functions in the following way: when a system carries out one of its functions we call this a process. For example, digestion is the process whereby the digestive system carries out its function. Notice also that a system is not always tied to a single process. Although the digestive system breaks down food, it uses different enzymes for different kinds of food. The digestive system breaks down lactose into its constituents (glucose and galactose) using the enzyme lactate, while it breaks sucrose into fructose and glucose by sucrase. Consider further that the same function can be multiply realized in one system. For example, the same system in plants carries out the light and Calvin cycle to produce ATP. Which process the system uses depends upon whether light is present (light cycle) or not (Calvin cycle).

Secondly, it is important to note the distinction between systems and processes: processes (digestion) are that which systems carry out. I understand a process to be a series of events, where an event is the loss or addition of a property. As some two-system advocates (notably Evans and Stanovich) separate processes and system in their most recent work, it is important to distinguish them. Furthermore, Evans and Stanovich think that two kinds of systems can perform the same kind of process. Stanovich is explicit on this point (see Stanovich 2011, ch 2), while Evans (2010a, ch 1) tacitly endorses this claim. I shall return to process individuation in chapter 3.

We can draw some further conclusions about system individuation from the above examples. There are at least three ways of individuating systems: functionally, structurally, and mereologically. I can be more precise with the following three ways of individuating systems:

> **Functional Individuation:** If x and y are systems that share all their functions then x and y are type identical systems, and if they differ in some function then they are not type identical (and so are not token identical either).[3]

Note that this allows that distinct token systems can share all their functions. Consider that my right and my left lungs have the same function but are distinct token systems. Instead, my right and left lung are distinct token systems for mereological reasons:

> **Mereological Individuation**: If x and y are mereologically distinct, then x and y are distinct token systems.

Mereological individuation says nothing about types. That two systems are not composed of the (numerically) same stuff implies nothing about whether the two systems are of the same kind. To be thorough, consider structural distinguishing of systems (understood as distinct from functional structure):

> **Structural Individuation**: If x and y differ structurally, then x and y are distinct token systems.

Again, structural individuation says nothing of type systems. To see why, consider two systems that are individuated solely on the basis of structure, such as cases of multiple realization. Consider, for example, a wine opener. There is the classic screw-and-single-lever wine opener, but there is also the screw-and-double-wing wine opener. Both serve the same function of opening wine bottles, but they differ structurally. Of course these two kinds of wine openers will be mereologically distinct. Are there cases of spatially coincident, functionally similar, structurally distinct systems? Perhaps two different operating systems both located on my hard drive (say Window and Linex) differ in the structure they use to operate my system. However, since they are both located on my hard drive, they are mereologically collocated. Furthermore, they serve the same general function, and thus are of the same system kind. They are distinct

[3] One might worry that functional individuation leads to a 'grain-problem.' The worry is that whether there is more than one kind of reasoning system simply depends on how 'fine-grained' we characterize the functions carried out by reasoning systems: we might regard a process as arising from two distinct systems, or as arising from one system that has two components. I will reply to this objection in the next chapter.

because of their structural differences.[4] One might think that Windows and Linex are not of the same kind functionally, since they differ in the ways that they operate. To what extended they do count as a case of multiple realization depends upon which account of multiple realization one adopts.

2.1. Multiple realizability, structural and functional individuation

In this section, I will examine three characterizations of multiple realizability: Putnam's, Shapiro's, and Aizawa and Gillett's. On Putnam and Aizawa and Gillett's accounts, structural and functional individuation come apart, while Shapiro's structural individuation collapeses into functional individuation. There are two related issues in the multiple realizability literature: 1) characterizing multiple realizability and 2) determining whether multiple realizability of the mental is compatible with type-type reductive physicalism. Here I am concerned only with the first issue. The existences of multiply realized properties (even radically multiply realized properties) became widely accepted through the influence of Putnam and (later) Fodor. Putnam defines functional isomorphism as follows: "Two systems are functionally isomorphic if there is a correspondence between the states of one and the states of the other that preserves functional relations" (1967). To put this more formally, any two systems A and B are functionally isomorphic just in case for any state x followed by state y in system A, state x is followed by state y in system B. The possibility of multiply realized systems follows immediately from Putnam's functional isomorphism. On this account of multiple realizability, there will be systems that are structurally distinguished rather than functionally distinguished.

[4] I should point out that distinguishing Structural Individuation from Functional Individuation does not commit me to the view that structural (or qualitative) properties are ontologically distinct from functional (or dispositional) properties. While it is plausible to think that there is no change in functional properties without a change in structural properties (i.e. that function supervenes on structure. See Prior, Pargetter, and Jackson 1982), many metaphysicians have argued for a stronger connection between quality and function: there is no ontological distinction between functional and structural properties. There is merely a conceptual distinction (See Martin 2008, Heil 2003, Mumford 1998, Mugg 2013). On either account, if two systems are to be functionally individuated, they will be structurally distinct as well. The latter view says something stronger than mere supervenience: two systems are functionally distinct if and only if they are structurally distinct. These debates are (to my mind) closely related to the question of whether there are genuine cases of multiple realizability. Again, I am not committed to the existence of multiply realizable properties or kinds. My taxonomy of system individuation allows those who think there are cases of multiple realizable properties or kinds a separate way of individuating systems. Those who deny multiple realizability will simply claim that, whenever there is a structural difference between two systems, there will be some functional difference as well.

Shapiro (2000), in defending type-type reduction from multiple realizability, does not deny that functionally isomorphic systems might be realizations of the same kind, but they would not be realizations of the same kind *in virtue of being functionally isomorphic*. He supports his claim with two examples. First, consider a mousetrap and a collection of cards with stages in a sequence (i.e. 'mouse enters trap,' 'trap activated,' 'mouse unable to leave trap,'…etc.). These systems are functionally isomorphic. For any state that the actual mousetrap is in, which is followed by a second state, the set of cards possess the same sequence. However, a set of cards with descriptions of mousetrap states is not of the same kind as the actual mousetrap. Thus, some systems are functionally isomorphic but are not realizations of the same kind. The mistake, Shapiro claims, is in identifying a description of sequences of relations as the system itself. Systems are not descriptions of sequences.

Shapiro claims that in order for there to be a case of genuine multiple realizability, there must be different ways to bring about the function that defines that particular kind. According to Shapiro, a property is multiply realized if there are distinct *functional* analyses of that property. For a property to be multiply realized, it must differ in its R-property at different times, where R-property is a label for "properties of realizations whose differences suffice to explain why the realizations of which they are properties count as different in kind" (2004, p. 52). This is why, according to Shapiro, different colored cork screws are not of different kinds—their causal properties do not differ with regard to their being cork screws. This leads to a dilemma for advocates of multiple realizability. Either realizing kinds differ in their causal structure, or they do not. Suppose that they do not differ in their causal structure. In that case, they are not different kinds at the lower level (such is the case of the differently colored cork screws). Suppose instead that the systems differ in their causal properties relevant to their being that kind (the wing cork screw and the lever cork screw). If the two systems differ in the causal properties that distinguish them as a certain type of system, then they are really distinct kinds of systems. Thus, this would not be a case of multiple realizability.

On this account of multiple realizability, structural individuation collapses into functional individuation. To see why, suppose X and Y carry out the same function, and are multiply realized. On Shapiro's account of multiple realizability, their being multiply realized entails that X and Y have distinct R-properties. If X and Y have distinct R-properties, then X and Y differ

functionally, at least at their realizer base. Thus, X and Y would be of the same kind at one level of consideration, but distinct kinds at a lower level. In the case or cognitive systems, suppose that two reasoning systems are multiply realized in Shapiro's sense. Then the two reasoning systems are of the same general kind (in that they carry out reasoning), but the properties that realize their being reasoning systems (their R-properties) differ. This is fine, but the debate between one and two-system theorists is whether S1 and S2 do in fact have different R-properties. Two-system theorists can agree that (at some level of functional analysis) S1 and S2 are of the same kind. After all, they are both reasoning systems.

Aizawa and Gillett (2009) offer an alternative to Shapiro's account of multiple realizability. Their somewhat baroque definition is worth quoting at length:

> "A property G is multiply realized if and only if (i) under conditions $, and individual s has an instance of property G in virtue of the powers contributed by instances of properties/relations F1-Fn to s, or s's constituents, but not vice versa; (ii) under conditions $* (which may or may not be identical to $), an individual s* (which may or may not be identical to s) has an instance of a property G in virtue of the powers contributed by instances of properties/relations F*1-F*m to s* or s*'s constituents, but not vice versa; (iii) F1-Fn≠ F*1-F*m and (iv) under conditions $ and $*, F1-Fn and F*1-F*m are at the same scientific level of properties." (p. 188)

Unsurprisingly, given Aizawa and Gillett's non-reductive physicalist commitments, this is a much more liberal characterization of multiple realizability. On this account, two cork screws of the same shape but made of different material (e.g. aluminum and steel) are a case of multiple realization. The extent to which F_1-F_n and F^*_1-F^*_m must differ is vague. Knoop hardness can be multiply realized in two objects with different kinds of material. However, they would not be multiply realized if the two objects differ only in that the first has one million molecules, but the second was one million and one molecules. One reason for disallowing this second case is that, if *any* difference between two objects with the same functional property implies that that property is multiply realized, then it would seem that *any* two non-qualitatively identical tokens of a type would be cases of multiple realization. In reply to this worry, Aizawa and Gillett say that "given the truly vast numbers of realizers involved at the microphysical level such a conclusion seems far from being either surprising or problematic" (p. 192).

On Aizawa and Gillett's account of multiple realizability, even if the functional properties of system X and system Y's realization base do not differ, X and Y may still be exemplars of multiple realization. In virtue of what are, X and Y *multiply* realized? They are

multiply realized because they differ in the structure of their realization base. So, on this account of multiple realizability, there will be systems that are structurally distinguished rather than functionally distinguished.

2.2. Individuating cognitive systems

I now turn to the implications for system individuation in cognitive systems. If (putative) two cognitive systems differ in their function, then they are not the same type or token system. One system might, however, have different processes for carrying out its function. The digestive system has a different enzyme for different molecules it needs to break down, and a reasoning system might have different ways of carrying out its function given different stimuli. A case of multiple processes carried out by one system (upon which the one and two-system advocates can agree) would be that of deductive and inductive reasoning. Inductive and deductive reasoning both count as reasoning and are carried out by the same system even though these reasoning processes differ. Furthermore, a single system can carry out some distinct processes simultaneously. The digestive system's various enzymes can break down different molecules at the same time.

Since I am not committing myself to the claim that functional differences are necessary for distinct token systems, it could turn out that there are multiple token reasoning systems of the same kind (like my right and left lung redundancy). Such a position is a theoretical possibility, and one that has not been explored in the literature. Perhaps this is due, in part, to the fact that, if the multiple systems perform the same function in the same way, then support for distinctness would require some evidence beyond difference in functionality. In the cognitive domain, this is difficult to do. Although we cannot individuate my lungs on functional individuation, we can on mereological individuation. However, mereological distinctness is not so simple when talking about cognitive systems. Interestingly, Evans claims that mereological distinctness is a necessary condition for two-system theory. He writes: "there must be neurologically distinct areas of the brain underlying" each of the two systems (Evans 2010a, p. 9 see also p. 44), and Goel provides some neurological evidence for the two-system theory. However, if one is not a localist about the brain (that is, if one thinks that generally a cognitive system will be located throughout the brain), then mereological individuation will be of little help in individuating cognitive systems.

Thus, it seems that Evans has set up too strong a criterion on system individuation, one which I will not require the two-system theorist to meet.

One might reply by citing that the circulatory system is distributed throughout a biological organism, but can be distinguished mereologically. Thus, the objection continues, denial of localization is consistent with mereological individuation. Thus, even if one is not a localist about the brain, S1 and S2 might be distinguished mereologically even though they are distributed throughout the brain. This argument is mistaken. We need to be careful about how we use the term 'distributed.' This argument mistakenly equates distributed with extended. It is not as though the circulatory system is 'gappy.' While my circulatory system is located in my right hand and left foot, those parts of the system are connected through other proper parts of the system. Those of us claiming that cognitive systems may be distributed throughout the brain will allow that these systems are gappy.

Here is a second objection. Classical mereology allows objects to be located across space. For example, David Lewis allows that the top half of a turkey and the bottom half of a trout (both currently connected to their respective halves) together compose an object. The existence of such 'mereological monsters' follows deductively from classical mereology's unrestricted composition. Perhaps cognitive systems could be like this: S2 might be partially located in the prefrontal cortex, but have proper parts located throughout the brain. While I do not deny that a system might be spatially gappy, we would need some reason to think that those parts compose a system. To do this requires that we first identify the system, leaving no work for mereological distinctness.

My point in offering these three ways of individuating cognitive systems is not to introduce new criteria, but rather to make the tacit criteria explicit. My ways of individuating cognitive systems should be relatively uncontroversial, at least among those who think the mind can be explained by dividing it into systems (Fodor 1968, 1983, Carruthers 2006, Pylyshyn 1984, Cummins 1975). Two-system theorists should be happy to agree with my characterization of system individuation because (I think) this is how they themselves have been thinking of system individuation.

2.3. Empirically distinguishing one and two-system theories

Given my characterization of systems, what would be the empirical difference between the one and two-system theories? A deceptively appealing strategy is system(s) failure. If there are two systems that operate independently of one another, then one system can fail while the other continues to operate normally. One way this might happen (though there may be others) is through a brain injury that might leave a patient's S2 impaired without altering his or her S1 or leaving their S1 impaired and not their S2. S1 can be partially inhibited, being a collection of systems, strictly speaking (according to all but Sloman). Suppose that there is a risk-determining system and a word-associative system, and both are parts of S1 (or are both Type-1 systems). Then a subject might be risk-determining deficient, but not word-associative deficient or vice versa. On the other hand, if there is just one system that operates in different modes, then that system either works or it does not. If the system is unable to bring the associative mode online, then the system is not functioning properly.

Determining whether double-dissociation is a good way to empirically distinguish one and two-system theories requires getting a bit more specific about which version of dual-process theory is at issue. I will argue that only on Sloman's account would S1 and S2 be double-dissociable. On default-interventionist accounts, S2 depends on S1 for its input. Thus, if S1 is taken offline, then S2 would have no input. Now, since S1 is typically thought of as a collection of module-like systems, S1 may be partially incapacitated. On this account, S1 may be partially incapacitated while leaving S2 intact. However, S2 will still lack the input from those Type-1 systems that are damaged, and so S2 may not function properly either. Interestingly, Evans does briefly argue for some dissociation when he considers the memory systems of the old and new mind. He explains that patient HM and other amnesic patients provide compelling evidence that there is a hippocampal learning system that belongs to S2. He also claims that the striatum or amygdala contain the S1 memory system because damage to these parts of the brain affect procedural, skill, and habit learning (Evans 2010a, p. 70). However, given the dependence the new mind bears to the old mind, it is unclear why this data favors his theory.

If parallel-competitive dual-system versions are right, we might be more optimistic about doubly-dissociating S1 from S2. However, even though Frankish and Carruthers's accounts are parallel-competitive, they run into similar problems, since S2 (being a virtual system) is realized

in the cycles of S1. S2 depends on S1. Thus, S2 cannot function without S1's functioning. Furthermore, S2 *just is* complex interactions between the module-like Type-1 systems of S1. So if the Type-1 systems are functioning properly, then S2 will be functioning as well. Sloman's account does seem to predict double-dissociation of S1 and S2, since S1 and S2 do not depend on one another and are both token systems. Interestingly, Sloman does not argue for his position on these grounds.

In addition to its being unclear that, on two-system theories, S1 and S2 are doubly-dissociable, it is also possible that the failure of a single system will be limited to a single mode and that the one-system might function only partially. If we think of the different modes of the reasoning system as dispositional properties of that system, then it is plausible that the system might lose one property while retaining others. Suppose that a subject fails to be able to reason associatively but continues to be able to reason in a rule-based way. The one-system theorist might claim that only the associative mode of the reasoning system is damaged. So whenever the subject reasons, it must reason non-associatively. It may be that the reasoning system is 'stuck' in its non-associative (i.e. rule-based) mode. That is, it would always be in its rule-based mode. Alternatively, the reasoning system might not be active at all when it should be operating associatively. Because of these different possible ways in which a system may malfunction, the one-system and two-system hypotheses will have similar explanatory power when it comes to failure of the reasoning system(s).

The one-system theory is an empirical claim and should be testable, but what would be the empirical difference between there being one reasoning system that operates in different modes and there being multiple systems? It will be important to contrast my one-system account against parallel-competitive dual-process theories separately from contrasting it with default-interventionist dual-process theories. In chapter 3, I will argue for a revised version of Sloman's (1996, 2002) Criterion S: simultaneous contradictory belief is incompatible with the one-system theory only if the reasoning processes resulting in those contradictory beliefs occurred simultaneously as well. Empirically distinguishing my account from default-interventionist dual-system theory will require that I fill in some details of my own account, and so I leave this task until chapter 5. In what follows, I merely attempt to undermine the inference to the best explanation for dual-system theory. I will explain how the one-system theory can account for the

data and point out how some evidence from the reasoning literature may actually fit better with the one-system theory.

3. Undercutting the inference to the best explanation for two-system theory

3.1. Wason Selection task

In the Wason Selection task, subjects are presented with four cards. Two are marked with a letter (say, A and D) and two with a number (say, 7 and 12). Subjects are told that each card has a number on one side and a letter on the other. Subjects are then asked to determine whether a material conditional (such as 'if there is a vowel on one side of the card, then the number on the other side is even') is true by flipping over the cards. They are also told to flip over as few cards as possible. Of course, a material conditional is only false when the antecedent is true and the consequent false. Thus, subjects ought to select cards with vowels (which almost all subjects do) and cards that are not even (which a minority of subject do). Oddly enough, most (60-75%) select cards with an even number, but if the consequent of a material conditional is true, then the truth or falsity of the antecedent is irrelevant to the truth of the conditional. Psychologists call this a 'matching bias' because subjects are biased towards the terms mentioned in the rules. In support of the matching bias interpretation, consider that if the rule is phrased as:

> If there is a G on one side of the card, then there is NOT a 4 on the other side of the card.

then a plurality of subjects pick the cards with 'G's and '4's, which is correct (Evans 1996). More specifically, the same number of subjects selected true antecedents (67%) in cases with and without a negation in the consequent, while the number of subjects selecting for a false consequent rose from 10% in the conditional without a negation in the consequent to 40% for the conditional with a negation in the consequent. These results are well confirmed.

While people are notoriously bad at reasoning about conditionals when the example is abstract, they are generally better (and perhaps even good) when the statement is drawn from a more realistic situation. For example, subjects do well on this task when the rule reads 'if you are drinking alcohol, then you must be over 18,' and the cards have ages or names of beverages printed on them. This is especially the case when the subject is told to imagine herself as an

officer who is supposed to determine whether the drinking rule offered above is being violated (see Griggs and Cox 1982).

Much ink has been spilled trying to explain subjects' performance on the Wason Selection task. There are at least two questions any explanation of this data must answer. First, why is it that most subjects exhibit a matching bias when the rule is given in the abstract, rather than getting it right when it is presented in the abstract? Second, why (given that most subjects are bad at the task in the abstract) are subjects good at this task when the rule is concrete? The two-system theorists claim that the best interpretation of this data is that there are two distinct systems of reasoning. That is, that there is one system operating in abstract cases (S2) and a different system (S1) operating for realistic cases. Since it is automatic, S1 is always operating. Thus, either S1 gets the response incorrect in the abstract case but gets it right in the concrete rule case (or S1 gets it wrong in both cases), but in the realistic case S2, overrides the S1 response.

There is a problem here for the two-system advocate. If S2 is a specialized abstract rule-based system, then why is it that subjects do poorly on these tasks? Two-system advocates claim that the mistake in reasoning is due to S1's dominance in this task. S1 is automatic, S2 is not. However, given enough time and effort S2 should come online. The task requires the use of rules, and it would seem as though that should activate S2. Evans (1984, 1989, 1996a) offers an explanation for why S2 is not activated in the Wason Selection task. Reasoning proceeds in two stages: heuristic and analytic. In the first stage, S1 decides what is relevant to solving the problem, and in the second stage, S2 performs a logical analysis of that relevant information. But all there is to completing the Wason Selection task is relevance: once a subject chooses which cards should be flipped (which cards are relevant), the task is completed. Thus, S2 is never activated. However, a further explanation is required for why it is that some subjects do activate their S2. Furthermore, Evans's explanation leaves out why it is that subjects do much better with concrete rules. I am not so much objecting to Evans's explanation, as I am pointing out that simply adopting the existence of the two distinct systems does not fully explain the data from the Wason Selection task.

The one-system theory is consistent with the above findings, and there are several ways in which the one-system theory might explain the above findings. To begin, notice that Evans's

explanation is compatible with there being one reasoning system. The one reasoning system could begin in a heuristic mode in which it determines what is relevant to answering the question at hand and then perform computations on that information. There is no need to locate these two steps into two systems, since a single system can carry out multi-step operations. One-system theorists can use Evans's insight that 'relevance' is all there is to the Wason Selection task. Once the single reasoning system determines which cards are relevant, the task is complete, and so the reasoning system does not use its rule-based operations.

Why is it, then, that some subjects do answer correctly? Perhaps the one-system theorist might claim that subjects who have been taught logic acquire a new protocol for determining what is relevant. Or perhaps subjects who answer correctly try to translate the abstract material conditional into a concrete and deontic example, whereas those who answer incorrectly do not. The single reasoning system proceeds in two steps when working properly. First, it translates the abstract example into a concrete example. For example, it might translate the material conditional 'if there is a vowel on one side of the card, then the number on the other side is even' into 'if they are drinking alcohol, then they have to be 18 or older' (substituting 'drinking alcohol' for vowel and '18 or older' for even number). Second, the system checks to see if the concrete deontic rule is violated. This is how many of us teach logic to our students. When confronted with a sentential syllogism, we tell students to think of a situation in which the premises are true and we see if they can make the conclusion false. That is, we ask them to mentally model the syllogism. Something similar might be going on in the Wason Selection task. Subjects are better or worse at the task depending upon how well they are able to substitute realistic content for the abstract content.

3.2. The attraction effect

Huber, Payne, and Puto (1982) developed an experiment that has been taken to support the two-system theory. Subjects were divided into two groups. In the first group, subjects faced a decision that was difficult because two options traded off in ways relevant to the subject's decision. For example, they had to decide between two six packs of beer with different costs and different quality ratings. The 'target' (the best choice) was given a quality rating of 50 and cost $1.80. The 'competitor' was given a quality rating of 70 but cost $2.60. In addition to the

original two options, the second group was given a 'decoy' option which was inferior to both of the original two options. Returning to the beer example, a decoy might cost $1.80 but only have a quality rating of 40. For this second group, a rational decision-maker would eliminate the decoy option and then face the same choice that the first group faced. Thus, the second group faces the same decision that the first group faces after they eliminate the decoy. If this is what subjects do (if subjects make decisions rationally), then we should expect that subjects in the second group would take longer to decide than the first group. After all, they must make exactly one more decision than the first group (namely eliminating the decoy). However, members of the second group reached a decision, on average, *faster* than members of the first group. This is known as the attraction effect.

Here is a way that the two-system theory can account for the discrepancy between what we would expect a rational decision-maker to do and what people actually do. In both cases, S2 is required to ensure that the right choice is made. However, it is also slow and uses cognitive energy. Therefore, a rational decision-maker would take longer to decide in the second case. Now, why is it that subjects do not do this? Humans will expend as little cognitive energy as possible. One way to put this is that humans are 'cognitive misers' (Stanovich 2011, Kahneman 2011); they think as little as possible. S2 requires the use of more cognitive energy than using only S1. When the decoy is added the subject might still recognize that careful rational (and S2) reasoning is required, but (thinking that such reasoning would be too hard) the subject 'gives up.' That is, S2 goes offline. It is not that S2 does not seem relevant, but that using S2 would be too cognitively costly to the individual. With S2 offline the subject is left to his or her automatic S1 response. Since S1 is a fast system, the decision is quickly reached.

Masicampo and Baumeister (2008) were interested in blood sugar effects on human reasoning, and glucose levels' effects on the attraction effect in particular. They ran a study demonstrating that the attraction effect increases when subjects' glucose levels are depleted. Masicampo and Baumeister claim that this supports the hypothesis that S1 is responsible for the attraction effect. However, this way of putting it is somewhat misleading, even assuming the two-system explanation is right. To account for the time difference in decision-making, two-system theorists need to say that the first group uses S2 while the second group uses S1 to complete the task. If both groups were using S1 to complete the task, then the time to decide

would be nearly identical between the two groups. The two-system theorist should claim that the limited resources of S2 are to blame for the attraction effect.

One-system theory is compatible with the attraction effect. That humans are cognitive misers does not depend on there being two-systems. So the one-system theorist might admit that subjects avoid engaging in difficult reasoning whenever possible. When there are many options, the reasoning system operates in its associative and fast mode to avoid having to think too hard. Alternatively, given the results from Masicampo and Baumeister (2008), the one-system theorist might claim the reasoning system needs sucrose to switch from its associative to rule-based mode. So the one-system theory has at least two explanations for the attraction effect, and so the attraction effect is compatible with there being only one reasoning system.

3.3. Belief bias

Subjects are more likely to claim an argument is valid if the conclusion is one they already believe. This phenomenon is known as belief bias. Consider one study performed by Evans, Barston, and Pollard (1983) in which subjects were asked to determine whether the conclusion followed from the premises. Table 2.1 gives the syllogisms and acceptance rates.

Table 2.1: Syllogisms and Acceptance Rates (from Evans, Barston, and Pollard 1983)

Syllogism	**Believable**	**Unbelievable**
Valid	No cigarettes are inexpensive. Some addictive things are inexpensive. Therefore, some addictive things are not cigarettes. Acceptance rate: **92%**	No addictive things are inexpensive. Some cigarettes are inexpensive. Therefore, some cigarettes are not addictive Acceptance rate: **46%**
Invalid	No addictive things are inexpensive. Some cigarettes are inexpensive. Therefore, some addictive things are not cigarettes. Acceptance rate: **92%**	No cigarettes are inexpensive. Some addictive things are inexpensive. Therefore, some cigarettes are not addictive. Acceptance rate: **8%**

Why is it that subjects are much more likely to accept that a conclusion follows from the premises if they already believe the conclusion? The two-system advocates claim that the best

explanation is that there is an associative system and a rule-based system. The associative system is responsible for the irrational belief bias. How exactly does this work? Default-interventionist two-system theorists will say that when a conclusion looks correct (such as in the 'believable' conclusions), the individual forgoes using the rule-based system. Parallel-competitive two-system theorists will claim that when a conclusion looks correct the associative system 'beats out' the rule-based system.

Notice that there are two cases in which logic and believability conflict. Namely, the case in which the conclusion is believable but the argument is not valid and the case in which the conclusion is not believable but the argument is valid. Call these 'conflict cases'. Evans and Curtis-Holmes (2005) demonstrate that belief bias increases in conflict cases when subjects are under a 10 second time constraint. Evans and Curtis-Holmes argue that this is exactly as the two-system theorist predicts. Because S1 is fast and S2 slow, when subjects are under a time constraint, they do not have time to complete the sequential operations particular to S2, and so they reason associatively.

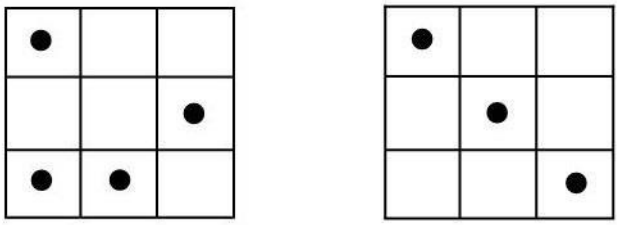
Two-system theorists also point to a series of experiments performed by Wim De Neys. A number of his experiments from a (2006) paper support the claim that belief bias increases with an increase on demands of working memory. Experiment 3 from his (2006) paper is worth explaining in detail. Subjects were divided into three groups based on how much working memory they had as demonstrated by the GOSPAN task.[5] Each of these groups were then divided into three groups: high interference, low interference, and no interference. Those in the high and low interference groups looked at a dot pattern and were told that they would have to remember the pattern. High interference subjects were given difficult patterns while the low group was given simple patterns (three dots in a diagonal row, for example (see Figure 2.1)). The last group was not asked to remember anything.

Belief bias in conflict cases increased with the increase of tax on their working memory, and this was true for subjects with high, medium, and low working memory resources. Those who had to memorize the dot pattern on the left performed more poorly on conflict cases than

[5] OSPAN stands for Operation Span Task, and GOSPAN is a group administrable adaptation of OSPAN. Basically, in this test, subjects must remember words that appear on the screen for a few seconds after the subjects compute simple arithmetic problems (e.g. $(2 * 2) + 1 = ?$). The more accurately subjects recall the word, the higher their working memory.

did those who had to memorize the box on the right, and both performed more poorly than those who did not have to memorize anything. Importantly, non-conflict cases (valid/believable and invalid/unbelievable) remained constant for all groups. De Neys claims that this supports the claim that all reasoners have access to two systems, and their ability to use S2 depends upon the availability of working memory resources. S1, on the other hand, is automatic and so does not involve "executive working memory" (p. 432).[6]

Figure 2.1: Example of High and Low Interference in De Neys (2006).
Example of what high interference subjects were asked to remember (left).
Example of what low interference subjects were asked to remember (right).



While belief bias increases under time constraints, there is a threshold after which there is no decrease in the belief bias effect. Two-system theorists need to account for this fact, and it does not seem to fit well with the two-system theory. According to the two-system theory the reason for belief bias is that S1 is faster than S2. S1 comes up with a response before S2 does, but if the reason for belief bias is that S1 operates more quickly than S2, then when subjects are given as much time as they would like, belief bias should decrease. Something more is responsible for belief bias. To put this in (two-system) theory laden terms: subjects' S2 is prone to errors as well. That there are two systems begins to explain belief bias, but an explanation of how S2 can fail to be fully rational is also required. This is not meant as an objection to the two-system explanation. That is, I am not claiming that there is data that counts against this explanation. Rather, I am pointing out that their explanation is incomplete. At best, there being two systems of reasoning is a partial explanation for the explanandum. More needs to be added for a complete explanation, and it is unclear how much work can be done in these details by

[6] De Neys and other two-system theorists assume that an automatic system does not use working memory. This is especially the case if, as two-system theorists maintain, automatic S1 systems are cognitively impenetrable and informationally encapsulated.

positing two systems. That is, further explanations may be of a kind of which the one-system theorist can help themselves.

The one-system theory can also begin to explain belief bias. One option would be to explain belief bias in terms of dispositionality, where dispositions can have degrees of strength (see Martin 2008 for support of the claim that dispositions come in degrees of strength). Subjects possess a disposition to claim that the argument is valid and a disposition to claim that the argument is invalid. When the premises are true it raises the strength of the disposition to claim that an argument is valid, and when the premises are false it raises the strength of the disposition to claim that an argument is invalid. These two opposing dispositions can be possessed by the same system, but both cannot *manifest* within the same system at the same time. That one system can possess opposing dispositions might seem odd, but consider a car stopped facing upward on a hill. It is disposed to move forward and backward at the same time. Which disposition manifests depends upon the stimuli obtaining. If the driver pushes on the gas while releasing the clutch, the car will go uphill. If the driver releases the break without pushing the gas and releasing the clutch, then the car will move downhill. If one system can possess opposing dispositions, then a single reasoning system can possess opposing dispositions. Of course, this option is open to the two-system theorist as well. My point is merely that there is no need to add a second system to our mental ontology if we explain belief bias with dispositions.

Since, on a two-system account, S1 always 'has its voice heard,' and its effects cannot be altogether escaped, we might expect that, for all responses to conflict cases, subjects will be more confident of their incorrect responses than their correct responses. Thus, the two-system theory predicts that subjects who respond incorrectly will (on average) be more confident than those who respond correctly. Indeed, Kahnman (2011) suggests that confidence is inversely correlated with normative correctness. However, subjects respond confidently in conflict cases even when they respond correctly. On the other hand, the one-system theory does not predict that subjects will be more confident when they deliver an incorrect response. There is empirical evidence against the two-system theory' prediction.

De Neys, Rossi, and Houdé (2013) were interested in whether there is internal conflict when subjects deliver an incorrect S1 response. Consider the bat and ball problem (Frederick 2005). Subjects are given the following:

> **Conflict Case:** A bat and ball together cost $1.10. The bat cost $1 more than the ball. How much does the ball cost?

80% of university students reply that the ball cost $0.10 (e.g., Bourgeois-Gironde and Vanderhenst 2009), but this is an incorrect answer. If the ball cost $0.10, then the bat costs $1.10 for a total of $1.20. The correct response is that the ball costs $0.05. (So the bat costs $1.05. So the bat and ball together cost $1.10). The bat and ball problem is a paradigmatic case of cognitive miserliness, and it is one of three problems that make up Frederick's Cognitive Reflection test (Frederick 2005). De Neys et. al gave subjects either the standard bat and ball case in which subjects are likely to deliver an incorrect heuristic response (except that they substituted pencil for bat and eraser for ball), or they gave subjects a version of the bat and ball problem that does not lead to an incorrect heuristic response, such as:[7]

> **No Conflict Case:** A magazine and a banana together cost $2.90. The magazine costs $2. How much does the banana cost?

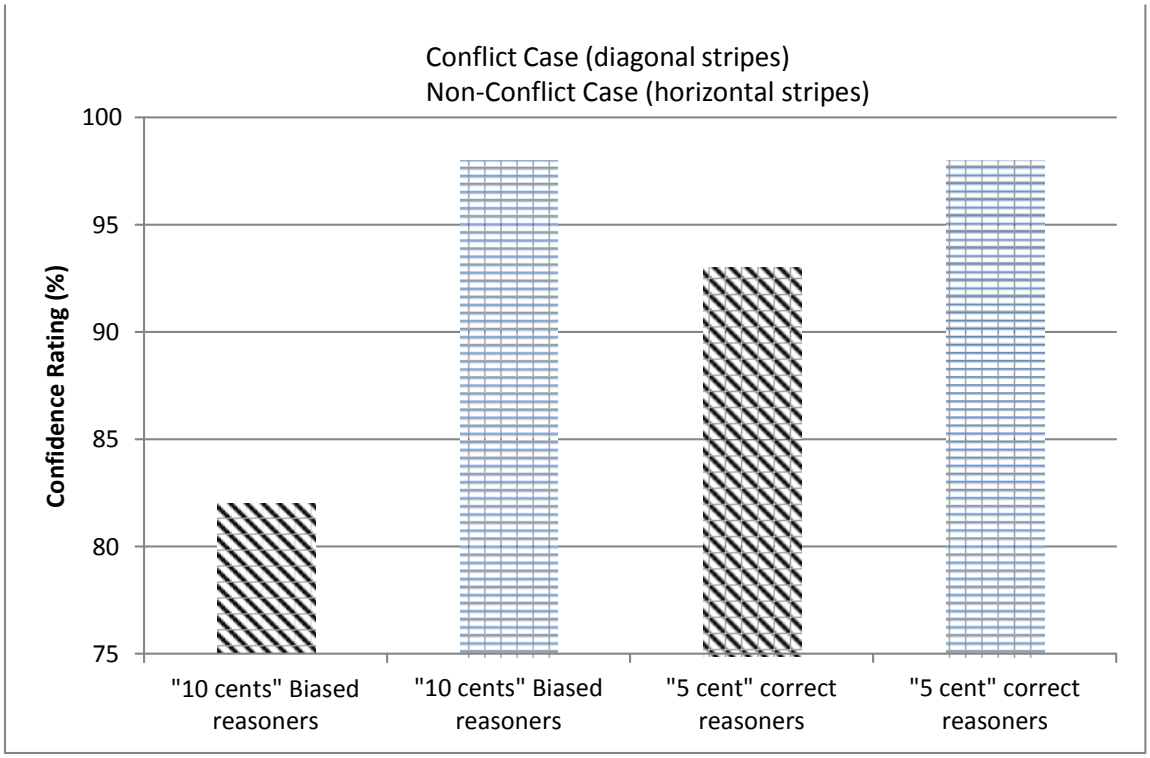
Subjects were then asked to report on a scale of 0-100 how confident they felt about their answer. Again, approximately 80% of subjects responded incorrectly in the typical bat and ball case. Below are the results (see Table 2.2).

Because confidence reports are themselves subject to biasing, De Neys et. al were not interested in the raw score given, but rather in the comparison between the conflict and non-conflict cases. Notice that subjects who responded incorrectly in the conflict case (diagonal bar on the left) reported lower confidence in their response that those who responded correctly (diagonal bar on the right). Furthermore, the difference in confidence between the conflict and non-conflict case was greater for the '10 cent' biased reasoners than for the '5 cent' correct reasoners. Thus, subjects who respond incorrectly in these difficult reasoning tasks do have some sense that there is something odd about the problem. Although De Neys and colleges do not say so, these results are contrary to the predictions made by the two-system theory. Let me be a bit more specific about why these results are problematic for the two-system theory.

---

[7] These versions of the bat and ball problem were translated into French, as these experiments were conducted in France.

Figure 2.2: Confidence in Conflict and Non-Conflict Cases



According to dual-process theory, the reason subjects tend to respond incorrectly in the bat and ball cases, *inter alia,* is because S2 does not come online. Recall that S2 governs metacognition. Indeed, on some accounts, S2 just is metacognition (see Carruthers 2009 and Frankish 2004). Thus, any recognition that one's intuitive response is incorrect (a metacognitive process) should come from an S2 process. Since subjects who get the bat and ball question wrong recognize (to some extent) that something is wrong with their response, their S2 processing is active. Thus, for subjects who get the bat and ball example wrong, their S2 both is and is not active.

Here is another way of posing the problem. Subjects who respond correctly in the conflict case continue to maintain their (incorrect) S1 response. Those delivering an S1 response are supposed to be 'blissfully ignorant.' Thus, those who deliver an S2 response should be less confident of their answer than those delivering an S1 response. Consequently, if the two-system theory is true, then in the bat and ball case, we should expect to see a higher confidence rating for those delivering an incorrect response than for those delivering a correct response. This is

exactly how Kahneman reasons about confidence and the two kinds of responses. For example, when he introduced a less intuitive process into his grading he became less confident of his responses.

> "I was now less happy with and less confident in my grades…but I recognized that this was a good sign, an indication that the new procedure was superior. The consistency I had enjoyed earlier was spurious; it produced a feeling of cognitive ease, and my System 2 was happy to lazily accept the final grade." (Kahneman 2011, p. 84).

Kahneman openly claims that S1 responses produce greater confidence than S2 responses. This is typical of two-system theorists, but inconsistent with the data from De Neys et. al (2013).

Of course, the two-system theorists can make adjustments elsewhere in their theory to accommodate these results. Namely, they can make alterations to their views about working memory. They might suggest that subjects have two separate working memory stores—one for S1 and one for S2. De Neys (himself a two-system theorist) has suggested that two-system theorists respond in this way. However, two-system theorists will then need to revise their explanation of belief bias increase under cognitive load. If S1 has a working memory of its own, why should belief bias increase under cognitive load? These questions are not unanswerable, but the results from De Neys et. al (2013) cannot be easily accommodated into the two-system framework. Problematically for some recent accounts (Stanovich 2011, Evans and Stanovich 2013a), Type-1 processes are defined as not involving working memory, and so they will be unable to avail themselves of De Neys's response that S1 (or Type-1 processing) possesses a working memory of its own.

On the other hand, the one-system theory can easily accommodate these findings. According to the one-system theory, subjects who respond correctly do not retain an incorrect response that remains active in reasoning. When subjects come to recognize the correct response, their incorrect response is erased. Because the incorrect response is erased (and subjects likely recognize that they have overcome some incorrect response in the bat and ball problem), they will be more confident than their incorrectly responding peers. Thus, according to the one-system theory, when subjects respond correctly, there should not be internal dissonance—even in the conflict cases. The data from De Neys et. al (2013) suggests that those who respond

incorrectly will experience dissonance rather than those who respond *correctly*, just as the one-system theory predicts.

3.4 Conjunction fallacy

Linda the bank-teller was devised by Tversky and Kahneman (1983). Subjects were given the following information:

> "Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations." (1983, p. 297)

Subjects were then asked which of the two following is more likely:

> A: Linda is a bank-teller.
> B: Linda is a bank-teller and is active in the feminist movement.

In some experimental trials, more than 80% of subjects said statement B was more likely than statement A. However, a conjunction can never be more likely than one of its conjuncts. A plausible explanation is that subjects reasoned by association. The Linda's description easily associates with Linda's being part of the feminist movement, but not her being a bank-teller. The description of Linda is more similar to that of a feminist bank-teller than it is to a stereotypical bank-teller. Sloman (along with Kahneman and Tversky 1983, Smith and Osherson 1989, and Shafir, Smith and Osherson, 1990) claims that because of the difference in similarity it is easier to imagine Linda being a bank-teller and active in the feminist movement than it is to imagine her as merely a bank-teller.

Two-system theory explains the data as follows. The associative reasoning system compares the two options by looking for properties associated with the description. Since being active in the feminist movement is associated with Linda's past history, S1 finds it easier to imagine a world in which Linda is a feminist than one in which she is not. For S1, the likelihood of objects possessing a property is directly proportional to how closely the concepts are associated. Thus, subjects claim that B is more likely than A. However, the rule-based system recognizes that A must be at least as strong as B because a conjunction can never be more likely than one of its conjuncts. Some subjects are able to use their S2 to determine this during the experiment, but most need to be prompted afterwards. Remember that S1 is automatic. It is

always online when the subject is reasoning. However, S2 is not automatic and requires effort when it comes online. Reasoners are cognitive misers; they will expend as little effort as they can. So if a subject can get away with only using his or her S1, he or she will. In the Linda case, it seems to subjects that they can get away with using only their S1, and subjects do not wait for S2 to complete its computation (or its response is overwhelmed by the S1 response) or (in default-interventionist versions) S2 will not even come online. A further explanation is needed for why it is that some subjects' S2 does come online, or why it is that some subjects' S2 is not overwhelmed by the S1 response. Perhaps subjects whose S2 does come online are less cognitively miserly, or perhaps subjects whose S2 comes online realize that S2 is needed to solve the problem. How is it that subjects recognize that S2 is needed? This amounts to the question of how S1 and S2 are regulated, a very important question to answer if a two-system account is to be adopted.

The one-system theory is also compatible with the Linda case. Perhaps the one reasoning system generally operates associatively for inductive arguments. For this reason, when subjects are asked to speculate about which statement is 'more likely' given a certain state of affairs, the system operates associatively. The 'more likely' phrase puts the system into an inductive mode. Of course, subjects know (implicitly) that a member of a conjunct is at least as probable as the conjunction of which it is a part. They may not put it exactly that way, but adult members of western societies seem to have the belief tacitly, since most subjects endorse this rule in exit interviews for the Linda case. Given this principle of probability (which will be true on any deductive or inductive logic), it follows deductively (by universal instantiation) that A is at least as probable as B. If there is only one reasoning system, then some subjects, when realizing that they can solve the question deductively, do exactly that. When they realize that they can solve the problem deductively, the system switches to deductive mode. So the Linda case is consistent with the hypothesis that there is one reasoning system that performs inductively (associatively) and deductively.

There is still a substantial difference between the one and two-system theories even though the one reasoning system can operate in two different modes of reasoning. Recall that the two-system theory is supposed to be a stronger thesis than the claim that we reason both deductively and inductively. The two-system advocates go to great lengths to explain that this is

not their view (i.e. that S1 is an inductive system and S2 is a deductive system). The inductive/deductive distinction is supposed to crosscut the S1/S2 distinction.

3.5. Argument strength

Similar examples to the Linda case can be found in experiments in which subjects are asked to gauge argument strength. Osherson, Smith, Wilkie, Lopez, and Shafir (1990) asked subjects to compare the strength of two arguments:

Argument 1
Robins have an ulnar artery.
Therefore, birds have an ulnar artery.

Argument 2
Robins have an ulnar artery.
Therefore, ostriches have an ulnar artery.

Most said that Argument 1 was stronger than Argument 2. Osherson et al. take this to indicate that subjects associate robins with birds more easily than they associate robins with ostriches.

Parallel two-system theorists will say that both reasoning systems receive the information that an object possesses a certain property. The output of the associative reasoning system is that objects closely associated with that object also possess that property, since the likelihood of an object possessing that property is closely related (and directly proportional) to how closely the concepts are associated. Thus, it is more likely that birds, of which 'robin' is an exemplar, possess the property in question than ostriches possessing that property. However, the rule-based system recognizes that the second argument must be at least as strong because ostriches are a type of bird, and so the conclusion of the first argument entails the conclusion of the second argument. The S2 of some subjects determines this during the experiment, but most need to be prompted afterwards.

Default-interventionist two-system theorists explain the data only slightly differently. Response to this problem begins in S1, where subjects form the (incorrect) associative response that Argument 1 is stronger than Argument 2. Subjects respond this way because they either 1) fail to recognize that S2 reasoning is required, 2) see that S2 reasoning is required but fail to bring S2 online, or 3) see that S2 reasoning is required, use S2, but S2's response is overwhelmed by the S1 response.

The one-system theory is also compatible with the above findings. My interpretation of the data is that subjects interpret the statement 'birds have an ulnar artery' as a generic claim—claims about kinds that are not reducible to claims quantified using 'all' or 'some.' If all the statements in Argument 1 and 2 are interpreted as generics, it seems plausible that Argument 1 is stronger than argument 2. Ostriches are odd birds, and many properties generally associated with birds (e.g. flight) do not apply to them. So perhaps robins' having an ulnar artery is more suggestive of *birds* having an ulnar artery than of *ostriches* having an ulnar artery.

There is good reason to think that subjects interpret the conclusions above as generics. Leslie argues that "generics articulate cognitively default generalizations" (2012, p. 33). This would explain why no known language has a word to signal generics, in contrast to quantifiers (all and some). Furthermore, there is good evidence that the first kind of generalization children acquire are generics. Hollander, Gelman, and Star (2002) asked English-speaking 3-year-olds, 4-year-olds, and adults a series of yes/no questions about kinds using existential, universal, and generic statements. They found that 3-year olds responded the same way regardless of the form of the question, and that there was no significant difference between 3-year-old, 4-year-old, and adult responses to generic statements. Tardif, Gelman, Fu, and Zhu (2010) replicated this study in Mandarin-speaking preschoolers, finding that 3-year-olds and 4-year-olds gave indistinguishable responses for universal, existential, and generic statements. This suggests that the default reading of a sentence is as a generic and that this is not an artifact of language, but that the effect is cognitive.

If subjects interpret the conclusions as generics, then ostriches not having an ulnar artery would not falsify the claim that birds have an ulnar artery. The proposition 'birds fly' is not falsified by the fact that ostriches do not fly. Leslie (2007, 2008, 2012) suggests that generics are sensitive to non-quantitative, content-based factors, especially the *most striking* characteristics of a kind. Leslie (2007, 2008, 2012) argues that generics are sensitive to non-quantitative content-based factors, especially the most striking characteristics of a kind. Subjects were told that a novel kind of animal, 'lorches,' sometimes have purple feathers, and that, in rare cases, these feathers are poisonous. Leslie (2007, 2008) found that subjects tended to affirm the claim 'lorches have poisonous feathers' even though most lorch feathers were not poisonous. This supports her hypothesis that generics are sensitive to striking characteristics of a kind. 'Birds

have an ulnar artery' interpreted as a generic has different truth conditions (tied closely to normalcy) than it would when interpreted as a universal. If subjects understand the conclusions in Arguments 1 and 2 as generics, and they believe ostriches are atypical, then it would follow that Argument 1 would be stronger than Argument 2. In short, we can explain the seemingly odd generalizations subjects make by assuming that subjects default to generic interpretations of claims about kinds.

The two-system theorists (specifically Sloman) have replied in two ways to proposals that the explanation involves generics. First, in response to the suggestion that Osherson et al.'s (1990) results are due to quantifier ambiguity, Sloman (1996, 2002) points out (in a footnote) that Shafir, Smith, & Osherson (1990) found inclusion fallacies using 'every single' wording, and Shafir et al. explained the meaning of the words 'every single' to subjects before testing. However Sloman's response is of limited use here. Shafir et al. (1990) found that subjects judged 'every single bank teller (in a particular bank) is conservative' to be more likely than 'every single feminist bank teller (in that bank) is conservative' (p. 243). This is a particularly problematic case, since "feminist conservative" might have been interpreted (at least in 1990) as a contradiction in terms, which would make the second statement analytically false. Although the first claim would still be more likely than the second (if quantifying over the same set of bank-tellers), the case is complicated by more than quantification. While Shafir et al.'s results provide evidence that subjects commit the inclusion fallacy even when the quantification is clear, it does not count as evidence against the claim that quantification ambiguity is to blame for incorrect responses in Osherson et al. (1990). Furthermore, Leslie has argued that adults sometimes interpret statements of the form "all X are Y" as generics (Leslie, 2012, p. 38-39). Leslie (2011) found that approximately half of subjects affirmed universally quantified minority statements (e.g. 'all ducks lay eggs'). Even when adults first admitted that 'male ducks do not lay eggs,' approximately 20% went on to affirm that 'all ducks lay eggs.' So there is good reason to think that some subjects in Shafir et al.'s study were interpreting the universally quantified statements as generics. At the very least, the study should be run again, carefully explaining to subjects the meaning of 'all,' with the arguments rephrased as:

Argument 1′
All Robins have an ulnar artery.

Therefore, all birds have an ulnar artery.

Argument 2′
All Robins have an ulnar artery.
Therefore, all ostriches have an ulnar artery.

Suppose that quantifier ambiguity is not to blame for the incorrect responses in Osherson et al. (1990). Suppose that subjects are interpreting Argument 1 and 2 as universal claims. What can the one-system advocate say? Again, the one reasoning system might generally operate associatively for inductive arguments, and so phrases like 'more likely' put the system into its associative mode. Of course, subjects know that if all Fs have Gs, then for any particular F, it has a G. Again, though subjects would not explicitly put it this way, subjects seem to endorse something like it in exit interviews. So there is evidence that they endorse it tacitly. Given this principle, it follows deductively (by universal instantiation) that Argument 2 is at least as strong as Argument 1. As I suggested above, if there is only one reasoning system, then some subjects, when realizing that they can solve the question deductively, do exactly that. When they realize that they can (and should) solve the problem deductively, the system switches to deductive mode. So the above data is consistent with the hypothesis that there is one reasoning system that performs inductively (associatively) and deductively. That we can reason both deductively and inductively is enough to explain the data from Osherson et al. (1990) even if quantifier ambiguity does not do the trick. There is no need here to invoke S1 and S2.

4. Objection and reply: virtues of these theories

Suppose that I am right in thinking that a one-system interpretation is possible for any evidence generally taken to support the two-system theory. This would not, on its own, entail that the inference to the best explanation argument for the two-system theory fails. First, I need to show that my one-system interpretations are not *ad hoc*. One might claim that the two-system theory offers a more unified explanation of the empirical data than a one-system alternative. Second, there are other virtues of theories to consider: unification with entrenched theories, predictive value, parsimony, and testability. I will offer the outline of a reply here, but I will deal with a number of these issues in subsequent chapters.

One-system explanations are not *ad hoc*. At least, they are no more *ad hoc* than two-system explanations. Given that there is a reasoning error, the two-system explanations appeal to either S2 failing to come online in time (or at all), or the S2 response being beat out by the S1 response. On the other hand, a one-system theory appeals to modes of operation within a single system. In cases like the Wason selection task, both the one-system and two-system theorist must account for individual differences. Stanovich (2011) has offered a taxonomy for categorizing individual differences, which mitigates worries about *ad hocery* for (at least his version of) the two-system theory. In chapter 5, I will offer a rival taxonomy for individual difference which is compatible with the one-system theory.

Two-system theorists have not specified what it would take to falsify the two-system theory, and Evans and Stanovich (2013a) have recently defended the claim that their theories do not have novel predictions because "falsifiable predictions occur not at the level of paradigm or metatheory—whether this debate is taking place—but rather in the instantiation of such a broad framework in task level models… It must be understood…that [dual-process] frameworks cannot be falsified by the failure of any specific instantiation or experimental finding. Only specific models tailored to the tasks can be refuted in that way" (2013b, p. 263). Let's interpret Evans and Stanovich charitably as denying a positivistic account of falsifiability and claiming that there need not be a 'critical test' for a theory to be viable. We might say, instead, that testability is always contrastive: experimental results favor one theory over another rather than falsifying a theory simpliciter (Sober 1999). If this is right, then dual-process theorists should be faulted for failing to consider how one-system opponents of their view might account for the data. What is needed are tests that would empirically distinguish one and two-system theories. In chapter 3, I will explain an empirical difference between the one-system and parallel two-system theories and offer a way to test my proposed one-system theory. In chapter 5, I will empirically distinguish my account from default-interventionist accounts.

The one-system theory is more parsimonious than two-system theory. The two-system theory posits two kinds of (and token) systems, each of which operate in different modes, while the one-system theory posits only one kind of system which operates in many modes. Thus, the one-system theory is more ontologically parsimonious, since it posits fewer entities and fewer kinds of entities. However, parsimony only comes to bear on choice of theory when all other

things are equal, and I am skeptical that all other things will be equal in our comparing one-system to two-system theories of reasoning.

Parsimony is more complicated if we consider dual-process theory, since it does not posit distinct *entities,* but rather distinct *kinds of processes.* I am sympathetic to thinking that parsimony applies to kinds as well as entities. Indeed, as Lewis (1973, p. 87) points out, it may exclusively apply to kinds rather than entities. Here one might claim that dual-process theorists win out in parsimony, since they posit only two kinds of processing that reasoning undergoes, whereas my account (which allows the properties on the Standard Menu to cross-cut one another) must admit many kinds of processes. Notice, however, that I do not posit the existence of many kinds of processing. I deny that reasoning processes divide into further sub-kinds altogether. However, the dual-process theorist might reply that I must multiply entities, because I admit that there are process that are (for example) rule-based and fast, whereas dual-process theorists do not. This may be right, but, as I pointed out before, parsimony is the least important virtue of a theory.

Two-system theories are not well entrenched. Two-system theorists are quick to point out that their theories are new, perhaps even radical. This is precisely why two-system theory is so interesting: if it is true, we will need to re-examine any theory that relies on the details of how humans reason. If the two-system theory is true, so much the worse for theories that do not cohere with the two-system theory, and we will have our work cut out for us as philosophers and psychologists to adjust our theories accordingly. While this possibility of new theoretical work is exciting, it relies on the assumption that two-system theories do not already fit with entrenched philosophical and psychological theories. On the other hand, the one-system theory is the default position, and so it coheres well with other theories. Indeed, two-system theorists suggest that much of our understanding of ourselves *depends* on the one-system theory. Thus, it is not as though we are comparing two theories on equal playing fields. One-system theory is far better entrenched than two-system theory—it is the default position.

Two-system theorists might reply that two-system or dual-process accounts have been posited in many other domains (learning, mind-reading, reasoning, decision-making, vision, and social-psychology). However, it is unclear that these various dual-process theories are gesturing at the same two systems or processes. It is better to construe dual-process theorists across

domains as having a similar strategy for explaning data, rather than thinking that they are all pointing to the same two kinds of processes or converging on the same two systems. Indeed, Evans and Stanovich (2013a, 2013b) claim that there is no generic version of dual-process theory. If that is so, then the various dual-process accounts across domains does not point to entrenchment of dual-process theory, since the various theorists are, in fact, positing different systems.

5. Conclusion

Two-system theorists tend to support their theory by an inference to the best explanation. I have examined some paradigmatic experiments taken to support the two-system theory. I have argued that my alternative to the two-system theory can explain this data as well. Furthermore, two-system theory lacks other virtues of a theory—virtues that must be taken into consideration when making an inference to the best explanation. Thus, the inference to the best explanation for the two-system theory fails.

Chapter 3: The Simultaneous Contradictory Belief Constraint

In the last chapter, I undercut the inference to the best explanation for the distinctness and kind claims. In a widely cited paper, Steven Sloman (1996, 2002) offers a different kind of argument for the distinctness and kind claims, based on the existences of simultaneous contradictory belief (henceforth SCB). Although Sloman's paper is widely cited, SCB has not featured prominently in dual-process theories (with Frankish and Sloman as the exceptions). Perhaps this is due, in part, to the fact that it is unclear whether default-interventionist dual-process theory is compatible with the existence of SCB. I have three aims in this chapter. First, I aim to empirically distinguish the one-system theory from parallel dual-process theory. I do this by arguing that any one-system alternative to dual-system theory should reject the possibility of SCB arising from simultaneously operating reasoning processes. Second, I argue that Sloman's putative examples of SCB are not SCB. In general, they fail on grounds of *simultaneity* or *contradictoriness.* However, this need not result in a stalemate between parallel dual-process theory and one-system theory. I propose an experimental paradigm which utilizes executive functioning and would provide compelling evidence for the claim that contradictory beliefs are held *simultaneously*.

1. Why simultaneous contradictory belief?

According to Functional Individuation, kinds of systems are individuated by their carrying out distinct functions (see chapter 2 of this dissertation). That is, systems are individuated by the kinds of processes which they carry out. Thus, process type individuation is prior (ontologically) to system type individuation. For this reason, in this chapter I will speak of processes instead of systems. Steven Sloman (1996, 2002) claims that SCB is incompatible with there being only one reasoning system, a claim he takes to be tautological (personal correspondence). To illustrate how SCB implies multiple systems or processes, consider the Müller-Lyer illusion, where two lines of equal length appear to be different lengths—even after subjects have measured the lines. Subjects seem to possess the contradictory beliefs that 1) the lines are of different length and 2) the lines are of equal length. Since there is SCB, we can infer

that there are two token processes: *perception* delivers the belief that the lines are of different lengths, while *reasoning* delivers the belief that the lines are of equal lengths. Sloman claim that, similarly, subjects sometimes possess SCB arising from reasoning. Thus, there must be (at least) two token reasoning processes, and since Sloman assumes that these process must operate on different kinds of systems, there must be distinct (token and type) reasoning systems.[1]

More specifically, Sloman claims that a one-system account is incompatible with SCB. Sloman claims that 'Criterion S' is incompatible with a one-system account of reasoning:

> **Criterion S:** "A reasoning problem satisfies *Criterion S* if it causes people to simultaneously believe two contradictory responses" (1996, p. 11).

However, there being one reasoning system is compatible with there being multiple 'belief boxes' (in Fodor's 1975 sense), or with a subject's having implicit beliefs that conflict with their occurrent beliefs (given distinct ways of individuating these mental states). So the existence of SCB on its own does not imply that there are two token reasoning processes. However, Criterion S does gesture in the right direction. I will reformulate and defend Criterion S in terms of processing (rather than systems) in order to remain agnostic as to whether there must be distinct systems underlying distinct processes. Call this revised claim the Simultaneous Contradictory Belief Constraint on Reasoning Processes, (henceforth, SCB Constraint):

> **The SCB Constraint:** A token reasoning process cannot have two sub-processes operating simultaneously which result in simultaneous contradictory beliefs.

Since the SCB must be *produced* simultaneously by *reasoning processes,* the SCB Constraint more narrowly specifies what is incompatible with a one-system theory.

Before arguing for the SCB Constraint, I must clarify some terms. Again, I take a process to be a series of events. Importantly, processes are divisible. The process of a mousetrap being sprung can be divided into its constitutive events: the mouse moves the bait, the hold-down is released, and the hammer falls. Processes are also decomposable into sub-processes: the springing of the mousetrap might be a sub-process of the process of exterminating all the mice in

[1] One might interpret Frankish (2004, 2009, 2012) as offering similar evidence for dual-system (and in his case, two-mind) theory, since he has long supported a dual-attitude account of belief.

my apartment. Since processes are composed of events, processes are not infinitely divisible. Call a sub-process that cannot be further divided into sub-processes, a 'simple process.' I will assume that processes (and sub-processes) can be divided naturally (as opposed to merely conventionally). Consider the psychological process whereby a subject comes to have an afterimage. The subject stares at a green, black, and yellow American flag and then looks at a white wall. Because of the effects of the first image on the subject's cones, the after image will be the familiar red, white, and blue American flag. We include each of the events or sub-processes in this process for explanatory reasons because they bear causal relations to the other events or sub-processes. We exclude certain events and sub-processes from this process (such as the movement of the subject's hands) because they bear no relation of explanatory interest. Finally, instead of providing necessary and sufficient conditions for 'reasoning,' in this chapter I will only consider cases that are uncontroversial reasoning processes. This is Sloman's approach, and Samuels (2009) suggests that the best we can do is rely on our intuitive notion of 'reasoning' (see chapter 5 for an extended discussion on 'reasoning').[2]

One might worry that my assumptions about processes and their division gives rise to a version of the grain-problem (Atkinson & Wheeler 2004): respiration and blood circulation are distinct processes, but, the objection continues, perhaps they are merely sub-processes of the (very coarse-grained) process of maintaining bodily survival. There are two related worries: first, how to determine whether two token processes are in fact sub-processes of a coarser-grained token process (call this the 'token grain-problem'); second, how to determine whether two types of processes are in fact sub-processes of a coarser-grained type of process (call this the 'type grain-problem'). (I raised this later concern regarding system individuation in the previous chapter.) My SCB constraint gives us one principled way to determine whether two token processes are part of a more general reasoning process or not: simultaneously operating reasoning processes resulting in SCB are not part of a token reasoning process. The type grain-problem is a problem for dual-process and one-system theories alike. If it cannot be resolved, then the dialectic between dual-process and one-system accounts of reasoning is not substantive. However, the type-grain problem only arises if one-system theorists claim that the one reasoning

[2] Note that the more we include in 'reasoning,' the more likely it is that we will need two kinds of processes to account for it. However, if we draw the boundary too large, then dual-process theory becomes far less interesting.

system carries out multiple kinds of processing simultaneously, and this is exactly what one-system theorists deny. On pain of triviality, one-system and dual-process theorists alike should reject the claim that a token reasoning process can have two token sub-processes of distinct types which result in contradictory beliefs.

Sloman understands 'belief' broadly to mean "a propensity, feeling, or conviction that a response is appropriate even if it is not strong enough to be acted on" (2002, p. 384). This definition is too broad: one who has a vague feeling that p should not be said to believe that p (see Marcus 1990, Davidson 1984, Baker 1995, and Schwitzgebel 2001, 2002 for further discussion). Indeed, he himself does not always seem to abide by it, since in his analysis of the Müller-Lyer illusion he wavers on whether subjects' feeling that the lines are of different lengths implies belief that the lines are different lengths. [3] The more broadly we define belief, the more likely that the SCB Constraint will be met, and on some accounts of belief, it will be exceeding unlikely (or impossible) for there to be SCB. For my purposes here, minimally, beliefs must be capable of being in contradiction with other beliefs and be capable of combining to form new beliefs. The most common way of satisfying both conditions is to take beliefs to be attitudes towards propositions, and so, for the purposes of this dissertation, I will assume that beliefs are propositional attitudes.[4]

One objection to the SCB Constraint is that positing the existence of beliefs possessed by Type-1 processing commits a category mistake. Necessarily, beliefs are personal level entities: we ascribe beliefs to the organism as a whole rather than to cognitive systems or organs. Fulfilling the SCB Constraint requires the existence of beliefs at the Type-2 and Type-1 level (or S1 and S2 possessing beliefs). However, Type-1 processes are subpersonal, and as such it is a category mistake to attribute beliefs to them. Similarly, no cognitive *system* can possess beliefs

[3] While in his (1996) paper Sloman, seems to indicate that the Müller-Lyer supports the two-system theory about reasoning, in his (2002) he merely takes it to indicate that perception and knowledge are governed by distinct systems—a conclusion Sloman (2002) recognizes is consistent with the one-system theory. He explains that "the conclusion that two independent systems are at work depends critically on the fact that the *perception* and the knowledge are maintained simultaneously" (385, emphasis added). So the Müller-Lyer illusion supports the existence of two distinct systems, but not two distinct reasoning (or even cognitive) systems. Notice here that subjects do believe (in Sloman's sense) that the two lines are different lengths. It is for this reason that my SCB constraint requires that the two processes be *reasoning* processes. I understand reasoning as a kind of cognition, and cognition as distinct from perception. Thus, perception and reasoning processes are mutually exclusive.

[4] Some philosophers and psychologists have claimed that non-conceptual content can be 'contrary' (such as in the waterfall illusion). However, I leave this aside, since non-conceptual content is unlikely to play any important role in the dialectic between dual-process and one-system theory.

because beliefs are ascribed to the agent. Since meeting the SCB constraint requires a category mistake, the SCB Constraint could never be met.

It is misleading to say that, according to the dual-process theorist, beliefs are held at the Type-1 level, or to say that S1 or S2 *believe* anything. The claim is that Type-1 processing results in a belief that is held at the personal level, and this belief may be contradictory to a belief issuing from Type-2 reasoning, which is also held at the personal level. Likewise it is misleading to speak of Type-2 processing possessing beliefs, or even System 2 possessing beliefs, since System 2 is a system of the organism. It is true that, on dual-process accounts, personal-level reasoning will be Type-2 reasoning (see Frankish 2009). That is, any token process of reasoning at the personal level is, on dual-process accounts, identical to a token Type-2 reasoning process. However, this does not entail that Type-2 processing *possesses* the output (i.e. belief state) of that process. We cannot ascribe mental states to Type-1 or Type-2 processing or S1 or S2, but they do generate beliefs. Supposing that there are two distinct processes, the picture is that Type-1 and Type-2 processes (subpersonal and personal reasoning respectively) both issue a response, and these responses can be in contradiction with one another. However, both of these responses must be attributed to the organism as a whole, given that they are beliefs. That is, they are attributed at the personal level. While the processing may be subpersonal and personal, the resulting states are exclusively personal.

This leads to a second worry: isn't it true *a priori* that subjects cannot have SCB? Belief ascription is constrained by rationality, and (though perfect rationality is not required) surely the absence of SCB is required. If it is true *a priori* that subjects cannot have SCB, then the SCB Constraint could never be met, and it is therefore a nonstarter for empirical investigation of cognitive architecture.

I deny that the absence of SCB in a subject is necessary for that subject's possessing beliefs. We can act in ways that do not accord with that which we avow, and a plausible way to explain this phenomenon is to posit implicit beliefs that are incongruous with explicit beliefs. Again, subjects possess SCB in the case of the Müller-Lyer illusion. Now, it is true that, when I attribute belief P to a subject, I generally assume that the subject does not believe not P. However, that I attribute 'not believing not P' to a subject does not entail that the subject does not possess the belief not P. I am relying on a distinction between *attribution* and *possession* of

belief. Denying this distinction belies an anti-realist interpretationalism. On such an account of belief, my SCB Constraint may be a non-starter, but I am not troubled by this since I am skeptical that such interpretationalist accounts will give us a concept of belief suitable to aid in uncovering cognitive architecture. Although I deny that contradictory beliefs are produced by simultaneously operating reasoning processes, I do not deny the existence of SCB simpliciter (*a priori* or *a posteriori*).

My reply to the objections that the existence of SCB is a category mistake, or that it is *a priori* false that subjects possess SCB, leaves aside two-mind theorists (Evans, Frankish, and Stanovich). On Frankish's (2004) account, S1 is governed by the rationality constraint and S2 is not. It may still be misleading to say that each system 'possesses' belief though. On Frankish's account each system produces different kinds of mental states—different kinds of doxastic propositional attitudes—namely, belief and superbelief (see chapter 1 for further detail of Frankish's account). Supposing that possession of belief is sufficient for being a mind (see Frankish 2012, Carruthers 2013b, p. 243), if the two systems possess beliefs, they constitute two minds.[5] Interestingly, then, if the contradictory beliefs are *maintained* simultaneously *by distinct systems*, then those two distinct systems would constitute distinct minds. Thus, SCB arising from simultaneously operating processes and maintained by distinct systems would falsify the claim that humans possess only one mind.[6] Since I will argue that we have no good evidence for SCBs arising from simultaneously operating reasoning processes, I will not take up the issue of whether putative SCBs arising from simultaneously operating reasoning processes would be *maintained* by distinct systems (as opposed to *produced* by distinct systems).

With these assumptions in mind, I offer the following argument for the SCB Constraint: [7]

---

[5] While it is certainly true that if an *organism* possesses beliefs, then that *organism* has a mind, the stronger claim that if *anything* possesses beliefs, then that thing has a mind be suspect. One might object, for example, that my belief box possesses beliefs, but is not a mind. However, this is mistaken. My belief box does not 'possess' beliefs. For x to possess beliefs is for some possible attribution of beliefs to x to be true. However, it would be a category mistake to attribute a belief to my belief box.

[6] Given that Evans is skeptical of the possibility of SCB on his default-interventionist model (Evans and Stanovich 2013a), this casts serious worries the consistency of his holding a two-mind theory.

[7] As an interesting historical aside, Plato seems to adopt something along these lines in his argument for the division of the soul in the *Republic*. "It is obvious that the same thing will not be willing to do or undergo opposites in the same part of itself, in relation to the same thing, at the same time. So, if we ever find this happening in the soul, we'll know that we are not dealing with one thing but many" because it is not "possible for the same thing to stand still and move at the same time in the same part of itself." Rather, "one part of the person is standing still and

Premise 1: Suppose (for *reductio*) that a token reasoning process, x, can have two sub-processes operating simultaneously which result in SCB: Belief(p) and Belief(not-p).
Premise 2: A token simple process will not result in contradictory responses.
Conclusion 1: Therefore, there must be at least two token sub-processes of x: one with the output Belief(p), one with the output Belief(not-p). Call these sub-processes y and z respectively. (from P1 and P2).
Premise 3: If two simultaneously operating token processes result in SCB, then those processes differ functionally.
Conclusion 2: Therefore, y and z differ functionally. (from C1 and P3).
Premise 4: If two processes are functionally different, then they are different types of processes.
Conclusion 3: Therefore, y and z are different types of processes. (from C2 and P4).
Premise 5: A token reasoning process cannot have two token sub-processes of distinct types which result in contradictory beliefs.

But P5 and C3 imply a contradiction.

Conclusion 4: Therefore, a token reasoning process cannot have two sub-processes operating simultaneously which result in SCB. (Negation of P1)

More needs to be said in support of my premises. Since premise 1 is an assumption for *reductio*, I begin with premise 2. Remember that processes can be divided into sub-processes. If a token simple process delivers contradictory outputs (p and not-p), then that process can be divided into at least two sub-processes: one sub-process resulting in p, another resulting in not-p. Recall that a simple process is a process that cannot be further divided. Thus, the putative simple process which outputs contradictory responses is not a simple process.[8]

One might object that it is conceivable that a simple process produces Belief(p and not-p), but does so without first producing Belief(p) and Belief(not-p). In that case, the process cannot be neatly divided into a process that outputs p and a process that outputs not-p.[9] First, this objection begs the question, since it assumes that p and not-p were produced simultaneously. Second, there is too little detail in this objection. How does this putative process output Belief(p and not-p)? While conceptualization may be our guide to possibility, we must conceive in a

---

another part is moving" (436c, Grube translation). This is the principle Plato uses to argue that the soul is divided into reason, desire, and appetite.

[8] Because premise 2 concerns token simple processes (rather than types), it is irrelevant that two token simple processes of the same type may result in contradictory responses.

[9] Thanks to Devin Curry for this objection.

detailed way to check if there are contradictions in the imagined state of affairs. It is incumbent on the objector to produce a how-possible model. One might suggest that Spinozian accounts of belief formation offer such a how-possible model. Gilbert (1991) and Mandelbaum (2013) claim that, if a subject considers a proposition, that subject believes the proposition, though the subject may reject it shortly thereafter. On this account, subjects frequently possess SCB. As Mandelbaum puts it, whenever subjects entertain an absurd proposition (e.g. dogs are made of paper), they possess SCB because they also have a standing belief to the contrary (e.g. that dogs are not made of paper). However, notice that these two beliefs are produced at *different times* by *distinct processes*. Thus, Spinozian accounts of belief do not claim that a simple process may result in Belief(p and not-p).

In support of premise 3, suppose that two token processes (x and y) do not differ functionally. If x and y are perfectly functionally isomorphic (i.e. do not differ functionally), then for any given input a, x and y will deliver the same output. Thus, x and y will never deliver different outputs for the same input. Therefore, if x and y deliver different outputs for the same input, then they differ functionally. Thus, premise 4 is true. One might claim that this is too strict a law: perhaps x and y deliver different outputs due to a lapse in performance (rather than competence). However, a performance error in one process (but not the other) would still imply a functional difference.

Premise 4 is similar to the second conjunct of Functional Individuation, which I defended in the last chapter. There I argued:

> **Functional Individuation:** If x and y are systems that share all their functions then x and y are type identical systems, and if they differ in some function then they are not type identical (and so are not token identical either).

Although Premise 4 concerns processes rather than systems, I still take it to be relatively uncontroversial, at least among those who think the mind can be explained by dividing it into systems (Fodor 1968, 1983, Carruthers 2006, Pylyshyn 1984). A plausible way of individuating kinds of processes is to examine their functional similarities and differences. To return to an example from the previous chapter, we may say that digestion and circulation are distinct kinds of processes because they are functionally different. They serve different functions in the body,

and they differ in their own functional organization. So there is good reason to think that functionally different processes are different kinds of processes. One might object by claiming that processes should be individuated by the kinds of systems that carry them out. However, since I have argued that systems should be individuated functionally, even if kinds of processes are individuated by the kinds of systems that carry them out, process types will still be individuated functionally.

I have already argued for Premise 5 in my treatment of the grain-problem. One-system accounts of reasoning, should accept Premise 5 to avoid trivializing their view. One-system theorists do not (and should not) reply to dual-process theorists by pointing out that these distinct token processes of different kinds resulting in contradictory outputs are *really* sub-processes of a more general reasoning process.

I conclude that one-system accounts of reasoning are incompatible with there being a token reasoning process that issues SCB. In agreement with Sloman (1996), the *simultaneity* of the contradictory beliefs is essential to SCB's being evidence against a one-system alternative. The SCB Constraint gives us a clear way of distinguishing reasoning processes, and it empirically distinguishes one-system and parallel dual-process accounts of reasoning.[10]

The existence of SCB arising from simultaneous reasoning processes would favor parallel dual-process theory over one-system theory.[11] However, the existence of SCB is not necessary for the existence of two token reasoning systems. It is possible that the two token systems operate in much the same way as a singular system. My breathing might operate similarly if I had just one large lung rather than two token lungs. However, since dual-process and two-system theorists embrace the kind claim as well as the distinctness claim, parallel dual-process and two-system theorists should (and do) reject this possibility. However, parallel dual-process theorists need not claim that for *any* reasoning task, there will be SCB. Thus, while the

[10] Interestingly, since most two-system accounts say that S1 is a collection of systems, they will allow the possibility that S1 will generate simultaneous contradictory outputs. However, Sloman has never claimed that S1 is a collection of systems, so the fact that he does not consider the possibility of simultaneous contradictory beliefs in his 1996 or 2002 article is not surprising.

[11] Interestingly, the existence of SCB arising from reasoning processes might favor parallel dual-process theory over default-interventionist dual-process theory.

presence of SCB would falsify the one-system theory, its absence (in one particular case) would not falsify dual-process theory.[12]

I now turn to challenge Sloman's putative examples of SCB. In each case, I will argue either that the two beliefs in question are not contradictory or that there is no reason to think that the beliefs are maintained simultaneously. This should not be surprising. If we rely exclusively on explicit responses from subjects, it is exceedingly unlikely that they will outright assert 'P and not-P.' In response to this difficulty for gathering evidence of SCB, I develop an experiment that does not rely solely on explicit avowals of subjects.

2. Putative SCB

One might think it obvious that the SCB constraint is met because there are subjects who possess implicit prejudices that contradict their explicit beliefs (see Devine et al. 1989, Frankish 2012). However, for implicit prejudices to fulfill the SCB Constraint one would need to show that (1) implicit prejudice tendencies are beliefs and (2) simultaneous reasoning processes form these contradictory beliefs (see Mandelbaum 2013 and De Houwer 2014 for further discussion). As such, I will focus on Sloman's examples of SCB, arguing that they do not meet the SCB Constraint. While each case does involve reasoning, each case either fails to provide *contradictory* beliefs or fails to demonstrate that the contradictory beliefs are held *simultaneously*.

2.1. Category inclusion

Consider Sloman's (1998) experiment in which participants "tended to project properties from a superordinate category to a subordinate category only to the extent that the categories were similar" (2002, p. 387).[13] He supports this claim through participants' judgments about argument strength. Subjects were given the following argument:

---

[12] While it is true that one can hold onto any empirical claim so long as one is willing to revise elsewhere (Duhem 1906), given the SCB Constraint and my arguments for it, any amendments to the one-system theory to make it compatible with the existence of SCB would be problematic.

[13] I should point out that there is no mention of two-system theories or dual-process interpretations of the data in Sloman (1998). Sloman's (1998) interpretation of the data is entirely (I think) compatible with a one-system theory.

Argument 1
Fact: Every individual piece of electronic equipment exhibits magnetic picofluctuation.
Conclusion: Every individual piece of audio equipment exhibits magnetic picofluctuation.

Subjects were instructed to assume that the 'fact' was true and then asked to determine the strength of the argument on a scale of 0 to 1 (0 being not at all convincing and 1 being very convincing). Subjects were not told anything regarding the enthymeme that "all audio equipment is electronic," but after filling out a probability strength questionnaire they were asked to assess category inclusions such as "all audio equipment is electronic" (they were given three choices: yes, no, and maybe). The mean probability judgment of the subjects who affirmed the conclusion in Argument 1 was .89 (given that they affirmed that all audio equipment was electronic), but of course, if all audio equipment is electronic, then (assuming that the 'fact' above is true) a rational subject would give the conclusion a probability of 1. Sloman points out that, when the category in the conclusion was atypical of the category in the premise, the judgments were even lower. For the following argument (Argument 2) the mean probability judgment was .76 (among those who affirmed the claim that all kitchen appliances where electronic).

Argument 2
Fact: Every individual piece of electronic equipment exhibits magnetic picofluctuation.
Conclusion: Every individual kitchen appliance exhibits magnetic picofluctuation.

During debriefing interviews, subjects agreed that there was good reason to assign Argument 2 the maximum probability because of the category inclusion. However, subjects also thought that their lower probability assessments were also sensible, though "they inevitably failed to express why" (2002, p. 387). Sloman concludes that, after being shown the correct answer, they had an associative response (a probability less than 1) and a rule-based response (a probability of 1) held simultaneously. These two responses on the part of the subjects are contradictory; therefore, Sloman concludes that this is a case of SCB.

There is an alternative explanation of Sloman's data, according to which ,subjects do not have SCB. While Osman (2004) argues that there is no *simultaneity* of the contradictory beliefs in this experiment, I will argue that there is no *contradictory* belief. As Sloman notes, subjects asserted that their initial responses were sensible even after coming to believe that a different

response was appropriate. Sloman claims this is good evidence that these opposing beliefs are held simultaneously. Notice that there is an enthymeme in both arguments crucial to their validity. In Argument 1 it is "every individual piece of audio equipment is a piece of electronic equipment," and in Argument 2 it is "every individual kitchen appliance is a piece of electronic equipment." My suggestion is that subjects took the strength of the argument to depend on the probability that these enthymemes are true. While the enthymemes seem plausible, they are not *certain*. Would a microphone stand qualify as part of the sound equipment? Does my corkscrew or manual egg beater qualify as a kitchen appliance? Even if we answer negatively, the questions give us pause. My suggestion is that participants were more inclined to exclude the microphone stand from audio equipment than the manual egg beater from kitchen appliances, which would explain why subjects claimed Argument 1 was stronger than Argument 2.

When experimenters told subjects that there is good reason to give Arguments 1 and 2 the maximal probability, subjects changed their response because they then took the enthymeme as guaranteed to be true. Thus, subjects could rightly claim that their original answers were reasonable. Their original answers reflected the degree to which they thought the enthymeme was true and, consequently (given that the conclusion follows only if the enthymeme is true), the degree to which they thought the conclusion was true.

My interpretation of the data predicts that subjects will correctly judge that the argument is valid when subjects are explicitly told to assume the enthymeme. Further experimental results from Sloman (1998) support this prediction. To support his interpretation of the data (that subjects project properties from a category to a subset of that category only to the extent that the two are similar), Sloman conducted an experiment in which subjects were presented with the same arguments as above, except that experimenters made the enthymeme explicit (Sloman 1998, p. 18-19). For example, they were presented with:

> **Argument 1′**
> Fact: All electronic equipment exhibits magnetic picofluctuation.
> Fact: All audio equipment is electronic equipment.
> Conclusion: All audio equipment exhibits magnetic picofluctuation.

When all of the premises were explicit, subjects overwhelmingly gave the argument the maximal probability of being true. Excluding one subject who gave every argument of this form .9, the

mean judgment was .99, and twenty-three out of twenty-seven subjects gave each valid argument a likelihood of 1 (Sloman 1998, p. 19). While Sloman takes these results to indicate that adding the enthymeme corrects their previous mistake, I take these results to indicate that subjects were judging conclusions to be less probable than 1 because they were not taking the truth of the enthymeme to be *guaranteed*. While Sloman's interpretation is consistent with these findings, my alternative interpretation predicts them.[14]

Sloman has two responses. First, in briefly considering the alternative explanation I am advocating, Sloman points out that in his earlier experiments subjects also affirmed the enthymeme during the experiment (the enthymemes were mixed with the argument strength questions). This is consistent with my interpretation of the data. Subjects endorsed the enthymeme, but not as a *certainty*. Subjects were only able to choose from three options: yes, no, and maybe. Sloman claims that if subjects were not *certain* about the enthymeme they would have responded "maybe" rather than "yes" (Sloman, 1998, p. 24). However, I do not have to be certain a claim is true in order to believe it (otherwise I'm not sure I believe much of anything), and subjects likely respond in a similar way. Experimentalists might ask subjects to rate how likely the enthymemes are true on a scale of 0 to 1. If my alternative explanation is right, there should be a strong correlation between argument strength and endorsement of the enthymeme.

A second reply Sloman offers comes from a pilot experiment he ran in which subjects were asked to rate the likelihood of inclusions and then given the 'fact/ conclusion' arguments explained above. Subjects gave 58% of the category inclusions a judgment of 1. Two out of 18 subjects gave maximal probability to each of the arguments. Interestingly, the mean of subjects who affirmed that the conclusion followed was higher than Sloman previously found (.96), but he says that this is "significantly less than 1" (p. 25). Sloman concludes that more is going on than just uncertainty about the enthymeme.

The data Sloman offers from this experiment is not all that we need. What we need is to compare a single individual's judgments of 1) the likelihood of a proposition and 2) the strength of a deductively valid argument which uses that proposition. Furthermore, Sloman needs it to be the case that subjects are maintaining these contradictory beliefs simultaneously. Given that

[14] This experiment also rules out the possibility that subjects give a likelihood of less than 1 because they think that there may be an equivocation in the terms being used, an alternative hypothesis to both Sloman's and my interpretation.

subjects affirm that the conclusion is (basically) guaranteed when the enthymeme is explicit, the best explanation is that subjects are altering how likely the enthymeme is from time to time.

Another possible contributing factor is that subjects are more skeptical of the truth of propositions when evaluating arguments (Mercier and Sperber 2011). Imagine you tell me that all Xs are Ys, and I agree. But if you tell me that all Xs are Ys, point out that all Ys are Zs (which you know I believe), and then point out that, as a result, all Xs must be Zs, then I might be more reluctant to endorse that claim that all Xs are Ys. Subjects might be more skeptical of the truth of propositions when those propositions are being used in arguments. We need to control for the possibility that subjects are simply altering the likelihood of the enthymeme based on whether it is being used in an inference or not. We can control for this possibility by running the following experiment. Give subjects arguments in the following format:

> **Argument 1′′**
> Fact: All electronic equipment exhibits magnetic picofluctuation.
> Claim: All audio equipment is electronic equipment.
> Conclusion: All audio equipment exhibits magnetic picofluctuation.

The idea here is to make the second proposition (here labeled 'claim') explicit while not telling subjects to assume it. Subjects should report how likely the claim is and how likely that the conclusion follows. If I am right, then there should be a strong correlation between these two scores.

I conclude that there is a plausible alternative to Sloman's interpretation of these results, according to which subjects do not have two responses in mind at the same time; they have one response at any given time. This one response depends upon how strongly the subject endorses the premises. Thus, Sloman's (1998) study does not give us a case of SCB.

2.2. Argument strength

Next, Sloman claims Osherson, Smith, Wilkie, Lopez, & Shafir (1990) give us an example of SCB. In this study, subjects were given the following two arguments and asked to say which was stronger:

Argument 3
Robins have an ulnar artery.
Therefore, birds have an ulnar artery.

Argument 4
Robins have an ulnar artery.
Therefore, ostriches have an ulnar artery.

Most said that Argument 3 was stronger than Argument 4. Osherson et al. take this to indicate that subjects associate robins with birds more easily than they associate robins with ostriches. Sloman thinks that subjects believe the correct answer—that Argument 4 is at least as strong as Argument 3—because they know ostriches are a kind of bird, and thus the conclusion of Argument 3 implies the conclusion of Argument 4. However, because they associate robins with birds more easily than ostriches, they (simultaneously) believe that Argument 3 is stronger. Thus, Sloman concludes, subjects simultaneously believe that Argument 3 is stronger than Argument 4 and that Argument 4 is at least as strong as Argument 3 (which is a contradiction).

This case fails to give us SCB because the beliefs are not *contradictory.* In the previous chapter, I argued that these results are due to quantifier ambiguity. Suppose that I am right. If these results are due to quantifier ambiguity, then the conclusion of Argument 3 does not imply the conclusion of Argument 4, and no contradiction arises. Sloman might point out that subjects, during exit interviews, admit that Argument 4 must be at least as strong as Argument 3 while maintaining that their original claim was sensible (see Sloman 1993). Thus, they do have SCB. However, this response fails to give us *contradictory* beliefs. During exit interviews, I suggest that subjects admitted that Argument 3 is at least as strong as Argument 4 because they changed their interpretation of the statements from generics to universals. Subjects believed that Argument 3 is stronger than Argument 4 *on the generic reading*, and they believed that Argument 4 is at least as strong as Argument 3 *on the universal reading*. There is, therefore, no contradiction here, merely an ambiguity in the argument presented to subjects. We should not expect subjects to be able to articulate these distinctions in justifying why their initial responses were sensible. After all, English does not have a generic operator (Leslie, 2012, p. 32).

2.3. Syllogistic reasoning

Sloman offers a third putative example of SCB, this one from Revlin, Leirer, Yopp, & Yopp (1980). Subjects were given 16 syllogistic arguments (half valid and half invalid). Subjects

where then asked what followed (they were given 5 options). Here Sloman only concerns himself with the valid arguments. Specifically, he cites the following two examples:

Argument 5
No members of the ad-hoc committee are women.
Some U.S. senators are members of the ad-hoc committee.
Therefore:

a. All U.S. senators are women.
b. No U.S. senators are women.
c. Some U.S. senators are women.
d. Some U.S. senators are not women.
e. None of the above is proven.

Argument 6
No U.S. governors are members of the Harem Club.
Some Arabian sheiks are members of the Harem Club.
Therefore:

a. All Arabian sheiks are U.S. governors.
b. No Arabian sheiks are U.S. governors.
c. Some Arabian sheiks are U.S. governors.
d. Some Arabian sheiks are not U.S. governors.
e. None of the above is proven.

Sloman claims that these arguments have exactly the same structure. In fact, they do not, though they are very similar. Below are argument structures. Notice the difference between the first premises in each argument.

Structure of Argument 5
X=Member of the ad hoc committee
Y=Woman
Z=U.S. Senator

No Xs are Ys.
Some Zs are Xs.
Therefore:

a. All Zs are Ys.
b. No Zs are Ys.
c. Some Zs are Ys.
d. Some Zs are not Ys.
e. None of the above is proven.

Structure of Argument 6
X=U.S. governor
Y=Harem Club
Z=Arabian sheik

No Ys are Xs.
Some Zs are Xs.
Therefore:

a. All Zs are Ys.
b. No Zs are Ys.
c. Some Zs are Ys.
d. Some Zs are not Ys.
e. None of the above is proven.

83% responded correctly for Argument 5 (d. Some U.S. senators are not women) while only 67% of participants responded correctly to Argument 6 (d. Some Arabian sheiks are not U.S. governors). In Argument 5 the right conclusion accords with our standing beliefs while in Argument 6 our standing beliefs tell us that the stronger answer (b. No Arabian sheiks are U.S. governors) is true. Sloman concludes from this example that "empirical belief obtained fairly directly through associative memory can inhibit the response generated by psycho-logic" (2002, p. 389). However, Sloman has given us no reason to think that subjects *simultaneously* believe that, in Argument 6, b and d both follow and that only d follows. We gain no explanatory power from positing the existence of SCB.

2.4. Conjunction fallacy

Finally, regarding Tversky and Kahneman's (1983) Linda case (see chapter 2 of this disseration), Sloman claims that he "can trace through the probability argument and concede its validity, while sensing that a state of affairs that [he] can imagine much more easily has a greater chance of obtaining" (1996, 12). Sloman also invokes a widely cited quote from Gould (1992): "I know that the [conjunction] is least probable, yet a little homunculus in my head continues to jump up and down, shouting at me—'but she can't just be a bank teller: read the description'" (p. 469).

Sloman pointed out to five of his department colleagues the difference in temporal relation between the responses in the Müller-Lyer illusion and the Necker cube illusion. Namely, in the Müller-Lyer case, the illusion that the two lines are different lengths persists even after one knows they are the same length, while, in the Necker cube illusion, one is only able to recognize one square as the front face at any given time. Sloman then asked his colleagues whether their experience in the Linda case was analogous to the Müller-Lyer or Necker cube case (see figures 3.1 and 3.2 below). All five agreed that the Linda case was like the Müller-Lyer illusion. Sloman then asked his colleagues whether the "Monty Hall" case[15] was analogous to the Müller-Lyer or

[15] In 'Monty Hall,' a subject is told to pick one of three doors, behind one of which is a car, behind the others are nothing. After picking, the game show host opens one of the doors which the subject did not pick. The subject is

Necker cube illusion. All of them (including Sloman) thought it was analogous to the Necker cube illusion. In the Monty Hall case the contradictory beliefs are not held simultaneously, whereas for the Linda case they are.

A number of opponents of dual-process theory have argued that the *phenomenology* of simultaneity does not imply that the two beliefs are actually held simultaneously. Pashler (1994) has suggested that subjects momentarily forget one response or the other. Likewise, Osman (2004) points out that there is no empirical evidence that these beliefs are maintained simultaneously. There are cases in which phenomenology leads us astray. Karen and Schul (2009) suggest that the phenomenology behind the Linda case is like that of phi phenomenon, in which subjects perceive two stimuli (say a ball on the right side of the screen and the left side of the screen), one right after the other, resulting in the illusion that the ball is moving across the screen (see Steinmean, Pizlo, and Pizlo 2000). Our phenomenology tells us that the two stimuli are moving, but they are in fact standing still.

Figure 3.1: The Necker Cube Illusion (right), which (Sloman says) does not produce simultaneous contradictory responses.

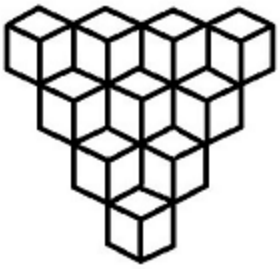
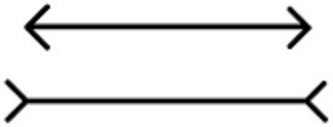


Figure 3.2: The Müller-Lyer Illusion (left), which does produce simultaneous contradictory responses.



We should not put much weight on the phenomenology of the theorists. Phenomenological evidence from naïve subjects (rather than experimenters) would be better; behavioral evidence would be better still. In order for phenomenological reports to count for anything here, we need more data. Sloman admits that his method was "highly unscientific" (2002, p. 386). Also, it is unclear whether 'sensing' constitutes belief, even implicit belief. Worse still, De Neys et al. (2013) found that subjects who respond incorrectly in cases like Linda

---

then asked if she would like to change her answer. While it seems that it does not matter, in fact there is a 2/3 chance that the car is behind the door that the subject had not picked initially.

experience the feeling that their response is not correct. Since these subjects do not have the correct belief issuing from Type-2 processing/S2, it is unclear that the phenomenology originates from SCB.

Sloman might support the importance of phenomenology by citing that, on his definition of belief, it is not necessary that subjects *act* upon their beliefs, and so a belief might not make a behavioral difference. Indeed, he claims that a belief can be a 'feeling' that a proposition is true. However, Sloman himself does not hold to this definition. In the Müller-Lyer illusion, he claims that subjects do not *believe* that the lines are different lengths after being shown that their lengths are equal, but subjects do 'feel' that they are different lengths. I recommend that we move away from phenomenological reports in favor of testable evidence for SCB.

3. Experiment to test for simultaneous contradictory belief

I have not argued that SCB does not exist; I have merely undercut Sloman's examples of SCB. A key difficulty is demonstrating *simultaneity* of contradictory beliefs (see Osman 2004). I will suggest a new paradigm for testing for the existence of SCB using measurements of cognitive depletion. According to the parallel versions of dual-process theory, the reason that subjects answer incorrectly in the examples above is that Type-1 processing beats out the slower Type-2 processes or that the Type-2 response is overwhelmed by the Type-1 response. Since Type-1 is automatic and "a person is aware only of the result of the computation," subjects who offer an incorrect answer arising from Type-1 processing, and then are told the correct answer, may keep the response generated by Type-1 processing (1996, p. 6, see also 2014, p. 77). If Type-1 processes are automatic and operate in parallel to Type-2 processes, then (for example) subjects continue to believe that Linda is more likely a feminist bank-teller than a bank-teller. Perhaps Type-1 processing, being automatic, is on-going in the presence of its stimulus, and, as such, continues to generate its output as the Type-2 processing is underway.[16] Subjects must suppress this incorrect belief if they are to maintain the correct answer while believing an incorrect, Type-1 generated, response. This suppression requires cognitive effort. Since

[16] This should assuage worries that, because Type-1 processing is faster than Type-2 processing, there will never be SCB on a dual-process theory (see De Neys 2013, Evans and Stanovich 2013). The reply that parallel dual-process theorists should make is that although Type-1 processing will end first, the subject 'hold off' in responding, and, in holding off, Type-2 processing has a chance to have its voice heard as well.

executive functioning is often likened to a muscle in that its use depletes it in the short term but strengthens it is the long term (see Muraven and Baumeister 2000), the suppression of the false belief should temporarily deplete the subject's cognitive resources. Therefore, those subjects who come to believe the correct answer after the reasoning task, through Type-2 processing, must expend more cognitive energy (and so be more cognitively tired) than those who remain naïve. With these theoretical points in place, I turn to how we might test for the existence of SCB.

In step one, subjects undertake a reasoning task that might involve SCB (e.g the Osherson et. al (1990) case). Subjects who offer the correct response should be dismissed. Those who offer the incorrect response should then be divided into two groups. The individuals are interviewed concerning the test that they just underwent. The first group will be made aware of their error during this interview. The experimenters will explain the category inclusion fallacy and apply it to the Osherson et. al (1990) case. The second group will not be made aware of their error. As a control for the possibility that it is the conversation about logic that is tiring, it is important that those in group two have a conversation that is as cognitively tiring as the conversation in the first. In the interviews with members of group two, experimenters will explain some unrelated fallacy that is supposed to be similar in phenomenology to the Necker cube (say, the Monty Hall example) (this is step two). Immediately following this interaction, members of both groups will be asked to complete a Stroop test, a typical measurement of executive functioning and cognitive depletion (step three). If subjects do possess SCB in this case, we should expect that the first group will have higher Stroop test interference (i.e. slower response times or less accurate responses) than the second group because they will have had to use their executive functioning to suppress their S1 belief that Argument 3' is stronger than Argument 4'.

One might claim that this test is unnecessary, since we already have evidence that belief bias increases under time constraints (e.g. Evans and Curtis-Holmes 2005) and under increased cognitive load (e.g. De Neys 2006). While both of these results are predicted by dual-process theories (though they are also predicted by one-system theory (see chapter 2, and Keren and Schul 2009), they have nothing to do with SCB (see also Osman 2004). My claim is not that, if belief bias increases under cognitive load, then subjects possess SCB. My claim is that, if

subjects possess SCB, then suppressing one of these beliefs should result in cognitively tiring these subjects.

4. Conclusion

Sloman offers the most explicit argument for two-system theory (or at least for the distinctiveness and kind claims). He argues that at least some of the experiments in the reasoning literature involve SCB, the explanation for which requires at least two reasoning processes. Sloman makes the further claim that these distinct kinds of processes are carried out by two token and type systems. I have argued that Sloman's putative examples of SCB fail, either on grounds that the beliefs are not contradictory (which could be easily controlled) or on grounds of simultaneity. However, the current lack of evidence does not imply a stalemate for advocates and skeptics of parallel dual-process theory. I have offered the outline for an experiment that would offer compelling evidence for the existence of SCB, thus strongly bolstering the claim that humans have at least two types of reasoning processing. This experiment also offers a clear way to test my one-system theory against parallel versions of dual-process and two-system theory.

Chapter 4: Rejecting the Cluster Kind Claim and Monothetic Kind Claim

In the previous two chapters, I argued that there is not convincing evidence for the distinctness and kind claims. Furthermore, a one-system alternative to two-system theory offers a better explanation of data from the heuristics and biases literature. Merely undercutting the evidence for a view may not be reason enough to reject the view, and a two-system theorist might suggest that even if humans possess only one reasoning system, it operates in two ways, along the lines of the Standard Menu.[1] In this chapter, I argue that even if humans have two kinds of reasoning systems, they are not divided in the way that two-system theorists claim. My arguments are equally problematic for dual-process theory, since dual-process theory includes the kind claim with regard to processes (rather than systems).

There are two ways to support the kind claim. Generally two-system theorists distinguish S1 type systems from S2 (and so identify the two kinds of systems) by the properties in the Standard Menu (see introduction chapter), and this is how dual-process theorists initially distinguished Type-1 from Type-2 processes (see Evans 2009a and Stanovich 2009). I called this the cluster kind claim. The second way to distinguish the two processes (or systems) is to choose a single property pair from the Standard Menu and use it as a necessary and sufficient condition for distinguishing two kinds of processing. Carruthers (2013a) and Evans and Stanovich (2013a, 2013b) have recently adopted this monothetic way of distinguishing reasoning processes (we might call this the monothetic kind claim). This chapter takes the form of a dilemma (the second horn of which includes a dilemma of its own). If the kind claim is true, then either the cluster kind claim or the monothetic kind claim is true. That is, either Type-1 and Type-2 processes (or S1 and S2) are distinguished using property clustering (i.e. the Standard Menu), or they are distinguished using a single property pair on the Standard Menu. They cannot be distinguished using property clustering from the Standard Menu because we have reason to think that these properties cross-cut each other in a way that is problematic for cluster kinds. The second horn of

[1] Evans's (2008, 2011b) shift to dual-process theory is consistent with the claim that there is only one reasoning system.

the dilemma will deal explicitly with two suggested monthetic accounts. Evans and Stanovich (2013a, 2013b) claim that a process is Type-1 if and only if it is autonomous whereas a process is Type-2 if and only if it is working memory involving. Carruthers (2013a) argues that there exists an intuitive system as opposed to a reflective system. I argue that these accounts face a dilemma: either they lack the explanatory power expected of kinds, or they tacitly commit themselves to cluster kinds along the lines of the Standard Menu. Thus, we should reject the monothetic kind claim as well.

1. The cluster kind claim

Although not every property on the Standard Menu is prevalent in every theorist's account of reasoning, some properties are present in every theory under consideration in this dissertation. The pairs of opposing properties that appear in every theory areas follows: fast/slow, heuristic (sometimes characterized as associative)/ rule-based[2], automatic/ controlled, and evolutionarily old/ evolutionarily new. While each theory might differ in which of these properties it chooses to emphasize, these properties remain constant across all theories under consideration in this dissertation.

In order to argue for (or against) a property cluster, two system theorists should take all instances of reasoning, determine what properties each process possesses, and then determine by statistical analysis whether there is a clustering. No such statistical analysis exists in the literature. Indeed, where the clustering is endorsed, it seems to be taken for granted. Here my aim will be more modest than such a statistical analysis. I will merely point out ways in which the various properties cross-cut[3] each other in *reasoning processes*.[4] There will be some cases in

[2] Carruthers is the one exception. He claims that S1 is computational, and therefore, rule-based rather than associative based. However, Carruthers and others who reject the claim that S1 is associative based will endorse the claim that it is heuristically based. It is for this reason that I offer the property distinction associative OR heuristic versus rule-based. Rule-based should be understood not as weakly computational (that it performs computations), but that it follows the normative rules of logic.

[3] For two pairs of properties to completely crosscut each other is for both members of one couplet to be able to combine with both members of the other couplet. More precisely, for any pair of properties A/B, A/B completely crosscuts some other pair of properties C/D just in case there exists objects w, x, y, z such that Aw, Cw, Ax, Dx, By, Cy, Bz, and Dz.

[4] One might wonder why I do not include cross-cutting examples from perception. I do not include such examples because the substantive claim being made by two-system theorists is that *reasoning* is divided into two kinds. Even if the properties on the Standard Menu cross-cut one another in perception, this is no problem for the two-system theorist.

which there is only partial cross-cutting, but in ways that are still problematic for the kind claim. It is possible that my examples are outliers, but then it will be incumbent on those using these properties for kindhood to argue that such clusters do exist.

Perhaps biological examples of cluster kinds will help. First, consider the two kinds of intestine: the small intestine is relatively long (~7 meters in length), with a relatively small diameter (3 cm), and absorbs high amounts of nutrients, while the large intestine is relatively short, with a relatively large diameter (6 cm), absorbs mainly water (rather than nutrients), and is primarily concerned with creating and excreting feces. These distinct properties are sufficient for the large and small intestine being distinct systems.[5] On the other hand, we do not consider the four sections of the colon (ascending, transvers, descending, and sigmoid) to be distinct systems, since they do not possess distinct property clusters—they all possess the properties outlined above, and can only be distinguished mereologically. Thus, they are not genuine distinct systems. Likewise with the putative kinds S1 and S2. If they are distinct kinds of systems, then they should be distinguishable by their distinct property clusters.

For the cross-cutting to be a problem, it needs to be the case that these properties cross-cut each other with regard to *reasoning* processes. Before arguing that each property pair cross-cuts each other property pair, I need to clarify the concepts that two-system theorists are using. Rather than attempting to define each concept, I will simply offer what I take to be relatively uncontroversial sufficient conditions.

1.1. Evolutionarily old/ evolutionarily new

That S1 evolved earlier than S2 is a recurring claim (Evans and Over 1996, Epstein and Pacini 1999, Reber 1993, Stanovich 1999). Sometimes evolutionarily new is a stand-in for 'that which only humans do' (see Evans 2003, 2009a). Evolutionarily old, on this account, is a stand-in for either 'that which humans and non-human animals do' or 'that which has long been thought to be shared with other animals.' Some putative examples of evolutionarily new

[5] One might claim that the small and large intestine are distinct kinds because they are mereologically distinct. However, that they are mereologically distinct does not, on its own, imply that they are distinct systems, since they are mereologically connected to one another. If we distinguished the two on the basis of mereology alone, we could divide the intestine into many more systems.

cognitive processes are: higher-order thought, hypothetical thinking, mind-reading, and language capacity. If a process involves one of these, then it is evolutionarily new.

1.2. Fast/slow

The process by which normal subjects generate an answer to the following question is fast: "what is 2+2?" An example of a slow process is where normal subjects compute "what is 17x24?" (The latter example is from Kahneman 2011, p. 20). The fast/slow distinction, in its modern incarnation, was introduced by Fodor in *The Modularity of Mind.* Given that S1 systems are supposed to be module-like, it seems fair to say that by *fast*, two-system theorists mean what Fodor means by the term. In Fodor's example of a fast process (speech shadowing), there is only 250 msec between the stimulus and response. In another example, Fodor cites Haber (1980), who showed subjects photographs for 10 seconds and tested them on their ability to correctly identify these pictures one hour later. Interestingly, performance of subjects "asymptotes at an exposure interval of about 2 seconds per slide" (Fodor 1983, p. 63). So perhaps it is fair to interpret Fodor as saying that a process is 'fast' when it is completed under 2 seconds or less, and this will work as a rough-and-ready benchmark for the purposes of operationalizing 'fast.'

Because processes are decomposable, there are two ways a process might be slow. First, processes might be slow because the process involves many steps. The example from Kahneman above is of this kind. Although it would take time to work out a response to the math problem, 17x24, each step in determining the answer might be fast. One might conclude that only complex processes (i.e. processes that decompose) are slow. However, a simple process (i.e. one that cannot decompose) might be slow. One might think that entertaining possible worlds to find a counterexample to categorical syllogism might be like this.

1.3. Associative (or heuristic)/rule-based

'Associative' and 'heuristic' are not interchangeable, but 'rule-based' is sometimes cited as the contrast for both of these properties. Heuristic processes are often construed as a kind of rule-based process, and so cannot be used as a contrast class with rule-based processes (see Gigerenzer and Regier 1996, Kruglanski and Gigerenzer 2011). A heuristic is a short-cut for solving a problem. All the specific explication of heuristic processes within the dual-process

literature meet this sufficient condition, whether it be heuristics as a 'rule of thumb' (Frankish 2004, p. 102), replacing a difficult question with an easy one (Kahneman and Frederick 2002), or the many instances of heuristics posited in the reasoning literature (representative heuristic, availability heuristic, etc.).

Rather than relying on a heuristic/ rule-based distinction, two-system theorists can, and often do, draw a contrast between associative processes and rule-based processes. Some two-system theorists rely heavily upon this distinction. For example, Kahneman (2011) calls S1 the 'associative machine' (but see also Sloman 1996, Morewedge and Kahneman 2011, Smith and DeCoster 2000, and Darlow and Sloman 2010). Associative processes can be divided into two temporal parts: formation and activation. In activation (or associative *thinking*), responses 'come more easily' to a subject given some precursor. For example, the word 'smoke' might activate the word 'fire.' Which responses come most naturally depends upon how concepts, words, or some other representations are associated through learning (formation). 'Fire' is activated by 'smoke' because I have seen these words, concepts, and things in conjunction many times. The associations made through learning should be the same as those made in reasoning. One might liken associative processes to a footpath: when one uses it, one strengthens it by packing down the dirt, making future use of that footpath easier.[6] As dual-process theorists use the terms, associations are supposed to be symmetric, while rule-based procedures allow for asymmetry. If A associates with B, then B associates with A. However, if A implies B, it does not follow that B implies A. Thus, if a process is asymmetrical, then it is not associative.

1.4. Automatic/controlled

Evans and Stanovich (2013a, 2013b) claim that Type-1 processes are "mandatory when their triggering stimuli are encountered and they are not dependent on input from high-level control systems" (p. 236). Evans and Stanovich call this property 'autonomy' (rather than 'automaticity') because these systems make minimal demands on working memory. However, notice how closely Evans and Stanovich's definition matches that of Fodor's (1983) definition of automaticity. Fodor claims that automatic processes are mandatory, in that when the stimulus is present, the process must be performed. This is a sufficient condition for being automatic. A

[6] Analogy taken from Mike Dacey.

second sufficient condition is the non-involvement of working memory. Controlled processes just are those that are not mandatory: if a process can be stopped when the stimulus conditions are present or if the process uses working memory, then it is controlled.

2. Examples of crossing-cutting properties

I now turn my attention to examples of how the properties on the Standard Menu cross-cut one another. There are six cases to consider. They are:

1. Evolutionarily old/ evolutionarily new and fast/slow.
2. Evolutionarily old/ evolutionarily new and associative/rule-based.
3. Evolutionarily old/ evolutionarily new and automatic/ controlled.
4. Associative/rule-based and fast/slow.
5. Associative/rule-based and controlled/automatic.
6. Controlled/ automatic and fast/slow.

2.1. The evolutionarily old/ evolutionarily new crosscuts the fast/slow

While two-system theorists have traditionally used the evolutionarily old/evolutionarily new distinction to distinguish the two systems (Evans and Over 1996, Reber 1993, Stanovich 1999), this claim has recently come under attack (Evans 2008, Samuels 2009, Caruthers 2013a, 2013b). In short, the worry is that some of the Type-1 systems in humans seem to involve very recent (and uniquely human) skills. Some systems that would typically be characterized as S1 involve language—which is evolutionarily new. Furthermore, belief bias—for which S1 is responsible—seems to arise from the prefrontal cortex, which is most developed in the human brain (Goel Buchel, Rith, and Olan 2000, Goel and Dolan 2003). This neurological data is at least some evidence that belief bias is evolutionarily new. Since belief bias is supposed to arise as a result of fast and automatic processing, this is evidence that there are process that are both evolutionarily new, fast, and automatic. This evidence is not decisive, since these processes might merely be piggybacking on evolutionarily new processes. Exactly how this might happen would depend on what one takes the relation between the two systems to be.

Humans use language, and presumably there is a system (or systems) responsible for comprehending language that is evolutionarily new. The language comprehension component of this system operates fast (in the sense described in the above section), as demonstrated in the psycholinguistic literature. Fodor's (1983) first example of a fast process is language

comprehension. Fodor (1983) offers experiments performed by Marslen-Wilson and Tyler (1981), who found that syllables are comprehended in approximately 250 ms, as evidence that language comprehension is fast. Marslen-Wilson and Tyler determined this through a shadowing effect in which a subject hears a word (usually through headphones) and repeats it out loud. Of course, the subjects repeating the words only implies that they are able to register and repeat the sounds quickly. Thus, one might claim that subjects do not *understand* what they are saying. However, there is good reason to think that subjects do understand the speech they are shadowing. Marslen-Wilson (1973) argues that subjects were able to correctly answer questions about content immediately after their shadowing a 300 word passage read out loud to them at 160 words per minute. Indeed, in the psycholiniguistic literature, researchers often assume that language *comprehension* is a fast process, and one task within the psycholinguistic literature has been to explain why language *production* takes a long time (relative to comprehension) (Griffen and Ferrera, 2006).

Furthermore, notice that the questions for which belief bias arise are language-involving (Evans, Barston, and Pollard 1983). Remember that belief bias increases when subjects are under short time constraints (Evans and Curtis-Holmes 2000). That belief bias increases under pressure indicates that the systems responsible for belief bias are fast: when there is little time to compute an answer, the system(s) responsible for belief bias do(es) so quickly. Thus, there are fast systems that have evolutionarily new components. Here the two-system theorist might reply that language comprehension systems are not Type-1 systems, but rather Type-1 processes recruit language comprehension systems. However, that some Type-1 systems recruit evolutionarily new language comprehension systems is evidence that those Type-1 systems are evolutionarily new; in order to operate, they rely on the resources of new cognitive systems.

In addition, there is evidence that evolutionarily old systems can be slow. Consider the experiments performed by Hare, Call, Agnetta, and Tomasello (2000) in which subordinate and dominant chimpanzees were placed on opposite sides of a room behind doors. In the middle of the room, in full view of both chimpanzees, was a desirable banana. Also in the middle of the room was a second banana with an opaque wall obstructing the view of this banana from the dominant chimpanzee's view but in full view of the subordinate chimpanzee. Researchers opened the two chimpanzees' cages approximately 15 centimetres for about 5 seconds, allowing

both to see the room, but not allowing them to exit the cage. Then the cage of the subordinate chimpanzee was opened just before the dominant's. Hare et al. found that the subordinate would only eat the banana that was out of view of the dominant chimpanzee. In a subsequent trial, the opaque barrier was replaced with a transparent barrier. In this trial the subordinate did not retrieve either banana. The two rival explanations for this data are the mind-reading hypothesis and the behavioral-dispositional explanation. Whether chimpanzees read minds or not, they do engage in complex behavior requiring time to work out what to do—that is, they engage in slow cognitive processes. It is precisely for this reason that Hare and colleagues would allow the subordinate and dominant to view the placement of the bananas for about 5 seconds before fully opening the doors.

One might reply that chimpanzees have an S2 and that S2 is engaged in performing this task. Perhaps this is the right response, though many two-system advocates have suggested (or outright claimed) that S2 is the system that only humans have. Even if we allow that problem-solving and planning performed by primates are carried out by S2, there are other examples in the animal kingdom of slow processes that would seem to be evolutionarily old. The Portia spider engages in a slow process by which it plans a route to its prey (Barrett 2011). It begins by looking at the prey and then (if no direct route is available) it begins scanning for broken or unbroken horizontal lines. If it finds a broken horizontal line, then it scans back toward the prey. If it does not find an unbroken horizontal line, then it continues to scan away from the prey. Once the spider discovers an unbroken horizontal line that leads from the prey to where it can go, it moves along that unbroken horizontal line to its prey. Carrying out this process takes time. So the Portia Spider's 'planning' of a route to its prey is a slow process, which at one time led scientists to believe that the spider was forming some kind of map in its very small brain (which we now have good reason to think is false). This process is slow, but is likely evolutionarily old.

One might object, claiming that it is not the cognitive *process* that is slow; rather, it is the *execution* of the process that is slow. An example to illustrate will be helpful. My examining a map and then following a mile long trail is cognitively simple. I need merely to remember what turns I must make at key points and to keep walking. The cognitive process of planning is fast, but it takes me a long time (in cognitive terms) to complete the mile long hike. The cognitive

process is fast, but the execution of the result of the processes is slow. This is unlike doing, say, a sentential logic derivation, where the cognitive process itself is slow.

The Portia spider case is more analogous to the logic derivation than it is to the hiking map-reading. The Portia spider uses its environment to plan the route and then executes the plan. It scans horizontally one way until it finds a break, then scans back the other way, and continues until it finds an unbroken horizontal line. This process looks similar to the process by which I work out an indirect derivation. I assume the negation of my conclusion, make all my premises explicit, and then spell out the consequences of those premises and my assumption until I find a contradiction. Both processes (rather than merely the execution) are themselves slow.

Whether a process is evolutionarily old or new is not a good indication of whether that process is going to be fast or slow. Non-human animals engage (at times) in slow thinking, and some paradigmatic Type-2 reasoning cases can be performed quickly.

2.2. The evolutionarily old/ evolutionarily new distinction crosscuts the associative/ rule-based distinction

There is evidence that some evolutionarily new systems operate associatively. Goel (2005) performed neuroimaging on subjects while they performed reasoning tasks that typically result in belief bias. Recall that it is the associative system that is supposed to be responsible for belief bias. Goel found activity in the prefrontal cortex, which is most developed in the mammalian brain (see also Goel et al. 2000, and Goel and Dolan 2003). This finding suggests that the system responsible for belief bias (an associative or heuristic system) is evolutionarily new. Evans suggests that a two-system theorist might reply that S1 and S2 will be implemented very differently in the advanced mammalian brain (Evans 2008). How exactly could the implementation of S1 differ in advanced mammalian brains? Perhaps Evans is suggesting that some Type-1 systems, although evolutionarily old, might have 'migrated' to an evolutionarily new part of the brain. Since functional distinctness is sufficient for system individuation, physical location of a system is not essential to system individuation.[7] Thus, a system might be

[7] Evans maintains that functional distinctness is not sufficient for individuation of the old and new mind (Evans 2010a). He claims that, in addition to functional differences, there must be distinct neurological correlates as well. I do not hold the two-system theorist to this high criterion: functional distinctness would be sufficient for individuation of Type-1 and Type-2 processing (see chapters 2 and 3 of this dissertation).

located in different parts of the mammalian brain than the non-mammalian brain. There does not seem to be anything incoherent about this suggestion, but it is *ad hoc* to posit that systems migrate to new regions of the brain in this manner. Furthermore, I think a simpler explanation is available: some evolutionarily new systems are associative or heuristic. It is plausible (though, as Evans has pointed out, not necessary) that evolutionarily old systems are physically located in evolutionarily old parts of the brain.[8] So Goel's evidence does suggest that there are evolutionarily new heuristic or associative systems.

Furthermore, language acquisition—an evolutionarily new process—seems to involve associations. While English-speaking children tend to master the past tense form of regular verbs in English by preschool (e.g. walk-walked), irregular verbs remain problematic for longer. There are about 180 irregular verbs in English that (it would seem) children must simply memorize. However, there are some phonologically related irregular verbs that follow patterns. For example, sing-sang and ring-rang. Interestingly, children often mistakingly extend this rule to words that phonologically associate (e.g. bring-brang). According to the Rumelhart-McClelland model, what is associated is sounds rather than words. In support of this claim, note that both children and adults sometimes generalize from (for example) fling-flung and cling-clung to spling-splung (Prasada and Pinker 1993, Xu and Pinker 1995). While Pinker and Ullman (2002) disagree with Rumelhart-McClelland's claim that *all* morphology is associative, they agree that "irregular verbs with overlapping partial similarity [are] best explained by the assumption that human memory is partly superpositional and associative" since solely rule-based theories must posit "needless complexity and esoteric representations, and fail to capture many linguistic, psychological, and neuropsychological phenomenon in which irregular forms behave like words" (p. 462).

What of rule-based evolutionarily old systems? This question can be answered in part by considering a related question: do non-human animals engage in rule-based procedures? Vervet monkeys seem to engage in rule-based inferences in the domain of social hierarchy. Vervets are able to recognize changes in social hierarchy and the implication of those changes for their position and the position of others within that hierarchy (see Cheney and Seyfarth 1985, and

[8] There is an epistemic reason for accepting the principle that old systems are physically located in old parts of the brain. As we move away from such a principle, we lose one principled way to tell whether a system is old or new ('one principled way,' not necessarily *all* principled ways).

Chase 1980 for similar data on rhesus monkeys). That one change in social hierarchy can alter the position of others in predictable ways, assuming a syntax for social relations, strongly suggests that this process is rule-based rather than associative. So it seems that there are evolutionarily old rule-based systems. The two-system theorists might bite the bullet and say that vervets have an S2. That would be a major concession, since first, vervets are not great apes, and second, this social reasoning system does not have any of the other typical features of S2.

One might object by saying that the process by which vervets reason about changes in social hierarchy is not rule-based, since, although vervets are very good at acting in accordance with the rule in one domain, they generally fail to reason correctly using the same rule in other domains. There is evidence that vervets can follow the transitivity rule in conspecific social settings, but there is no evidence that they follow the transitivity rule with other animals or with inanimate objects (Cheney and Seyfarth 1985). This objection depends on the following conditional: if subjects do not act in accordance with a rule in one domain, then they are not following the rule in any domain. I deny this conditional. Consider that, in the Wason Selection task, subjects perform well in deontolgical contexts but not indicative contexts. It would be odd to conclude that, even in the deontological context, subjects are not reasoning using rules. Thus, this objection fails.

I conclude that there are evolutionarily new processes that are associative and evolutionarily old processes that are rule-based.

2.3. The evolutionarily old/ evolutionarily new distinction crosscuts the automatic/ controlled distinction

The system responsible for language comprehension is a useful example of an automatic evolutionarily new system. Reading words comes automatically for literate individuals. When one sees the letters G R E E N together, we cannot help but reading them as ‘green.’ The Stroop Task is built on the assumption that literate people automatically read words. MacLeod (1991) gives an overview of work on the Stroop task generally, but the first section of his paper is useful for our purposes here. He points out that Stroop himself performed two separate experiments (which MacLeod replicated). In the first experiment, subjects were asked to read words referring to a color (‘color words’) from one of two decks. The first deck had color words written in

various ink colors. The second deck had color words written in black ink. The difference in speed at which the subjects read the two decks was not significant in Stroop's and MacLeod's tests (5.6% for Stroop, only .01% for MacLeod). In the second experiment subjects were asked to say what color of ink appeared on the card and were given one of two decks: a deck of ink blots, or a deck with color words written in various colors. Subjects identified the color of ink significantly faster for the first deck than the second (there was a 74% difference in speed in Stroop's experiment and 71% in MacLeod's). The upshot is that the mis-matched ink color and word names did not make a significant difference when subjects were reading, but it did make a difference when subjects were naming ink colors. One explanation for this is that reading of color words is automatic, whereas expressing ink color is not. Because reading color words is automatic and the language system is activated by seeing printed words, subjects cannot help but recognize the words as meaningful. So when subjects see R E D in green lettering, they automatically register the meaning 'red.' However, since color property itself is not automatically registered, it takes longer to register than the word meaning. If that is right, then language comprehension systems are automatic. Again, language comprehension systems are evolutionarily new, and so there are evolutionarily new automatic systems.

Additionally, there is good reason to think that some controlled processes are evolutionarily old. Macaques and capuchins are able to perform delay matching samples, which requires that the monkeys hold the images in mind for a brief period of time (see Miller, Erickson, Desimone 1996 for macaques, Tavares and Tomaz 2002 for capuchins). Or consider that apes in Köhler's (1927) famous experiments had to keep in their memory that there was a box in the opposite corner of the room which the animals could use to create a make-shift ladder in order to reach a treat hanging just out of reach. So it seems that there are evolutionarily old controlled systems.

I conclude that we have evidence that there are evolutionarily old processes that are controlled and that there are evolutionarily new automatic processes.

2.4. The associative/ rule-based distinction partially crosscuts the fast/slow distinction

There is reason to think that associative reasoning processes tend to be fast. Since associative activation occurs when some concept 'comes more easily' given a precursor, it would

seem that associative activation will be fast. As such, I will not offer an example of a slow associative process.

Rule-based processes can also be fast. For example, through practice, a rule-based procedure might become routine and thus be quickly processed. Consider the mathematician who quickly does long division in her head, or consider whether the following sentence contains an error:

Martina is from argentina.

I am sure you noticed that 'argentina' ought to be capitalized. To determine this you probably consulted the rules for capitalization, but the process was also quickly processed. (This second example is taken from Mallon and Nichols 2011). Thus, a rule-based process can be performed quickly. This should not be surprising because computations can be performed quickly, and computational processes are rule-based processes.

2.5. The associative/ rule-based distinction crosscuts the automatic/controlled distinction

Mallon and Nichols's example in the preceding section is of an automatic, rule-based process. We cannot help but see grammatical errors (of at least some kinds), and recognizing a grammatical error involves the use of a grammatical rule. Consider that when, marking papers, I might find a student who continually makes grammatical errors. Once I have pointed the student's errors out on the first pages, I might consciously decide to stop paying attention to grammatical mistakes in order to focus merely on content. However, as experience tells us, we are unable to stop recognizing the errors. For at least some errors, like those in a bad freshman paper, I do not have to focus on finding grammatical errors in order to spot them in a paper. Now, as Mallon and Nichols point out, recognition of grammatical error is a rule-based process. Grammatical rules are, at the risk of offering a triviality, rule-based. Thus, recognition of grammatical errors is (for at least some subjects) an automatic rule-based process.

There are also instances where we control associative processes. Two-system theorists claim that subjects reason associatively in cases like that of Linda the bank-teller and the robin case from Osherson et al. (1990). Do subjects reason automatically in these cases? The process would seem to involve very little effort. Is the process mandatory? It would not seem so. I do not

have to think that Linda is more likely a feminist bank-teller when I read the description. I can tune the description and problem out. I can refrain from thinking about the problem completely. However, it would seem that, once I begin to think about the implication of the description, I cannot help but think that Linda is more likely a feminist bank-teller. So the process is mandatory in the sense that it cannot be stopped once it is started, but not mandatory in the sense that one cannot help but do it. Thus, it is not mandatory as defined by Evans and Stanovich, following Fodor.

I have provided evidence that there are controlled associative reasoning processes, and that there are rule-based automatic processes.

2.6. The automatic/controlled distinction crosscuts the fast/slow distinction

Before considering empirical cases in which automatic processes are slow, I need to reply to some potential conceptual worries. One might think that it is necessarily true that whenever a process is automatic, that process is fast. After all, the way that researchers test for automaticity is generally based on speed (e.g. German and Cohen 2012). One might justify this practice by reasoning as follows: 1) if a process is automatic, then that process is highly efficient in that it does not use many cognitive resources, and 2) if a process is highly efficient, then it will be fast. Thus, an automatic process is a fast process. However, there is a clear conceptual distinction between automatic and fast. Efficiency does not necessarily imply speed. Thus, nothing in the definition of 'automatic' implies that all automatic processes are fast.

Consider that the process that begins when I touch a hot stove is automatic. First, the pain receptors in my hand send a signal to my brain. Second, I have a reflexive response in which I draw my hand back quickly and swear. All of this happens quickly, but not necessarily so. It is possible that my pain receptors might send the signal slowly to my brain, and my reflex to withdraw my hand might happen only after a few minutes. By definition, if the process must happen once the stimulus has been introduced, the process is automatic. In an episode of Warner Brother's children's show *Tiny Toon Adventures*, Elmira (a rather dull girl) picks up a scalding hot bottle. When the fluid leaves the bottle and hits the table, it eats through the wood like acid, and yet she does not notice immediately that it is hot. The narrator of this episode explains (with the help of a pseudo-medical diagram) that this is because Elmira's nervous system operates at a

fraction of the speed of a normal person's. But notice that the process is still automatic. Thus, there is no conceptual link between automaticity and speed.[9]

Slow automatic processes are not only possible, but are actual as well. Consider the Portia spider again. The spider's behavior would seem to be automatic. We can manipulate its behavior in certain predictable ways, which indicates that it follows set (perhaps innate) rules for scanning and capturing food. One might also think that the spider's planning is automatic, since we might think that the spider's brain is too simple to have control in a robust sense. Surely the spider does not have, for example, executive functioning.

Additionally, controlled processes can become fast over time. Controlled procedures might start out slow—my adding 2+2 for example—but, with time and practice, they can become fast. Similarly for logicians who know many logical systems. They can control which logical system they use (classical, dialetheism, or many-valued), and can control the cognitive processes by which they draw inferences using those logical systems, but after using the systems for a time, they are able to do so quickly.

An objection to the above examples is that a skill becomes fast because it becomes automatic over time. This will not help the two-system theorist. The fast/slow distinction admits degrees in a different way than the controlled/automatic distinction. Although one could define automaticity in such a way that it admits degrees (see Kunda 1999), the way dual-process theorists have defined automaticity does not allow for degrees: when the stimulus is present, the process is mandatory and does not required the use of working memory. A system might come to have less minimum requirements for activation, and so be active more easily, but this does not make it more automatic; it only changes the stimulus condition. On the other hand, the fast/slow distinction admits of smooth degrees. The upshot is that a process, although it may become faster, cannot slowly become more automatic. Therefore, this objection fails.

[9] One might object that we revoke certain conceptual links in certain types of fiction, such as cartoons. What is important for establishing possibility is to conceive in a maximal way, and cartoons require that we not conceive a state of affairs in a maximal way. For example, we must not think of the implications of a teacup's being able to talk, reason, and experience pain. As it happens, this cartoon helps us conceive in a maximal way.

3. The kind claim: Monothetic kinds

Recently, in response to the claim that these properties cross-cut one another (Keren and Schul 2009, Evans 2006, 2008, Carruthers 2013a), some prominent two-system theorists have abandoned the System 1/ System 2 distinction and adopted monothetic[10] dual-process accounts. Carruthers (2013a) maintains a two-system theory that distinguishes intuitive from reflective reasoning, which, he argues, is different than the received S1/S2 distinction. According to Carruthers, intuitive processes are unconscious, while reflective processes are conscious. Stanovich (2009, 2011), Evans (2008, 2009a), and Stanovich and Evans (2013a, 2013b) have abandoned two-system theory, but maintain that reasoning processes are of two kinds: Type-1 processes are autonomous and, so they claim, do not use working memory, whereas Type-2 processes require working memory because they involve cognitive decoupling and mental simulation. 'Type-1' and 'Type-2' are, on this account, natural kind terms. Because both Evans and Stanovich abandon the claim that these kinds of processes must be carried out by distinct kinds of systems, their views are a species of dual-process theory rather than two-system theory.

Although their accounts differ in important respects (see chapter 1 of this dissertation), Carruthers, Evans, and Stanovich's dual-process theories share similar deficiencies stemming from the fact that each are monothetic: for each theorist, the singular pair of properties that are supposed to establish natural kinds fails to do so. In this section, I argue that Evans, Stanovich, and Carruthers's accounts of natural cognitive kinds face two problems. First, their accounts allow too much to count as reasoning, as Sloman (2014) briefly notes. Second, each theory faces a dilemma: either the singular property that is necessary and sufficient for being a Type-1 (or intuitive) process cannot accomplish the explanatory work needed to support dual-process theory, which undercuts the theory as an account of natural kinds (since natural kinds should be explanatorily powerful, projectable, and used in prediction), or the account tacitly uses the various properties from the Standard Menu, thereby committing that account to the Standard View. I conclude that Evans and Stanovich's recent move from two-system theory to dual-

[10] I use the term 'monothetic' as opposed to 'essential' because, while Evans, Stanovich, and Carruthers all offer necessary and sufficient conditions for the two kinds of reasoning, 'essences' have typically been taken to be modal in nature as well. I will not hold these theorists to the claim that Type-1 or intuitive processing *could not* have been otherwise.

process theory and Carruthers's move from the S1/S2 distinction to an intuitive/reflective distinction do not succeed as defenses of the dual-process thesis.

3.1. Rejecting the standard menu

It is important to note that Evans, Stanovich, and Carruthers are all unsatisfied with the Standard View for a similar reason: namely the properties on the Standard Menu cross-cut one another. Although I have argued above for cases of cross-cutting, one might object to my following criticisms by claiming that Evans, Stanovich, and Carruthers admit cross-cutting of different kinds. Indeed, Stanovich and Evans maintain some very limited clustering. As such, it will be important to survey, briefly, the ways in which these theorists admit that the properties on the Standard Menu cross-cut one another.

I begin with Evans (2008, 2006). He claims that there is evidence of "a distinction between stimulus-bound and higher-order control process in many higher animals (Toates 2006), including rodents" (2008, p. 258). Furthermore, it is implausible that there is a common S1 to all animals "with a single evolutionary history" (p. 259). Evans (2006) argues that there is good reason to think that many Type-1 systems are evolutionarily new (p. 202). The processes responsible for belief bias are "certainly not 'ancient' in origin" even though they have other Type-1 features (p. 203). Importantly, Evans also rejects the characterization of S1 as associative, since "theories that contrast heuristic with analytic or systematic processing (Chen & Chaiken 1999, Evans 2006) seem to be talking about something different from associative processing" (2008, p. 261). He also says it is unwise to characterize S2 as rule-based "if only because it implies that S1 cognition does not involve rules" (2006, p. 204). The associative/rule-based distinction cross-cuts other important properties on the Standard Menu, since "rules can be concrete as well as abstract and any automatic cognitive system that can be modeled computationally can in some sense be described as following rules" (2006, p. 206, see also Gigerenzer & Regier 1996). Evans also says that S2 should not be characterized as "abstract and decontextualized" since these do not correlate with "slow, sequential, explicit, and rule-based" (2008, p. 261). While being conscious and being controlled are both associated with Type-1 processing, Evans points out that it is "far from clear" to what extent "conscious thinking really is 'in control' of behavior," and unconscious cognition can be intentional (2006, p. 204). Thus,

the "automatic-controlled distinction between the Systems 1 and 2 is far from clear cut" (p. 204) and "fraught with difficulties" (p. 206).

Carruthers (2013a) offers several cases of cross-cutting both to establish that S1 and S2 (as divided by the Standard Menu) are not natural kinds and to distinguish his own intuitive/reflective distinction from the S1/S2 distinction. First, heuristics can be rational, and are almost always ecologically rational (Gigerenzer, Todd, & ABC Research Group, 1999). Indeed, intuitive reasoning is often better than reflective reasoning, since decisions made using intuitive reasoning may lead to greater satisfaction with the outcome (Wilson et al. 1993). Thus, Carruthers denies his previous claim that heuristics are "quick and dirty" (2009, p. 110). Furthermore, reflective reasoning may employ heuristics (p. 16). Next, Carruthers (2013a) claims that intuitive reasoning can be slow, as when subjects use the "sleeping on it" heuristic or when subjects gain information about their partner's immune system through saliva obtained through kissing (Barrett, Dunbar, & Lyceett, 2002). Furthermore, Dijksterhuis, Bos, & van Baaren (2006) found that, for some reasoning tasks, subjects' intuitive responses conform to norms better than their reflective reasoning, and their work suggests (Carruthers argues) that unconscious reasoning may be slow. Carruthers also points to work in animal reasoning literature which suggests that rats and pigeons can track randomly "changing rates about as closely as is theoretically possible to do" (Carruthers, 2013a, p. 14) (Gallistel and Gibbon, 2001, Balci, Freestone, and Gallistel 2009). Carruthers argues that non-human animals engage in unreflective processes that can be flexible and rule-governed (p. 6). Importantly, he does so by claiming that intuitive processes are not associative, and so must be rule-based (Gallistel and Gibbon 2001, and Gallistel and King 2009).

Other theorists sympathetic to dual-process theory have noted cross-cutting examples as well. For example, Mallon & Nichols (2011) note that rule-based processes may be fast, as in the spotting of grammatical errors. (For cross-cutting examples from critics of dual-process theory See Keren & Schul, 2009, Kruglanski and Gigerenzer, 2011). It should give us pause that prominent dual-process theorists, who at one time used the Standard Menu to distinguish kinds, have rejected the Standard Menu. Now, since clustering does not require perfect correlation, dual-process theorists might maintain the S1/S2 distinction using the Standard Menu by arguing that these examples are mere outliers to an otherwise genuine correlation. However, given these

numerous examples, it is incumbent on advocates of the S1/S2 distinction to argue, contra Evans, Stanovich, and Carruthers, that these properties do cluster.

3.2. Two or three new proposals, and the stone soup objection

Evans and Stanovich moved away from the Standard View independently, but for similar reasons. Stanovich says that the S1/S2 distinction is problematic for two reasons: first it implies that there is just one S1, when in fact there is a set of module-like systems, which he calls The Autonomous Set of Systems (or TASS) (2009, 2011). While he has recently been more explicit on this point, even in his 1999 monograph, he claimed that S1 was a set of systems. Second, Stanovich admits that the properties on the Standard Menu do cross-cut one another. In response, he writes that "the defining feature of Type-1 processing is its autonomy—the execution of Type-1 processes is mandatory when their triggering stimuli are encountered, and they do not depend on input from high-level control systems" (2011, p.19). However, he goes on to say that some properties from the Standard Menu will closely correlate with autonomous processes: they will be fast, will not use much executive functioning or central processing, and will be able to operate in parallel, but these properties are not *essential* for a process to be Type-1.[11] The defining feature of Type-1 processing is autonomy.

Evans began to talk of processes rather than systems in his 2008 literature review of dual-process theory. In response to the examples of cross-cutting he provides, he suggested moving to a distinction between processes rather than systems, "since all theorists seem to contrast fast, automatic or unconscious processes with those that are slow, effortful, and conscious" (2008 p. 270, see also Evans 2009a). The move from *system* kinds to *process* kinds is less significant to the dialectic than it might appear. If the properties on the Standard Menu cross-cut one another in ways such that they cannot distinguish natural *system* kinds, then those same properties (which do not cluster) cannot distinguish natural *process* kinds either. Dual-process and dual-system theories are both theories about what natural cognitive kinds exist. The former claims that there are two kinds of *processing*, while the later claims that there are two kinds of *systems*. If the properties on the Standard Menu do not cluster, then the set of non-clustering properties cannot

[11] The only example of cross-cutting that would be a problem for this small cluster, given the cross-cutting cases outlined above, is Carruthers's claim that unconscious processing (i.e. that which he calls 'intuitive') can be slow.

be used to identify kinds of processes or systems. Thus, given that these properties do cross-cut one another, there is not a distinction to be made, using the properties on the Standard Menu, between kinds of systems or processes.[12] Evans (2011) latter claimed that the real distinction between Type-1 and Type-2 processes is autonomy/working memory involving, just as Stanovich did. Evans (2009a) defines autonomy as those processes "that can control behavior directly without need for any kind of controlled attention" (p. 42).[13] While Evans and Stanovich differ in important respects (i.e. they disagree on the distinctness claim), they agree on how to divide Type-1 and Type-2 processes (Evans and Stanovich 2013a).

Carruthers has recently argued that, while the property clusters on the Standard Menu do not mark out natural kinds, there is a real distinction between intuitive and reflective systems. That is, 'intuitive' and 'reflective' are natural kind terms designating kinds of systems. Intuitive and reflective systems are systems whose processing is unconscious or conscious respectively. Again, contra the Standard View, Carruthers argues that unconscious processes can be slow, controlled, and conform to the highest normative standards (2013a, p. 2-3), and conscious processes can employ heuristics and do not necessarily lead to improvement.

What is the relation between these proposals? Both claim that reasoning processes are of two kinds. For this reason alone, we may call both dual-process theories of reasoning. Both also agree that using the properties on the Standard Menu to identify and distinguish the two kinds is hopeless because the properties on the Standard Menu cross-cut one another. Finally, they agree that there is a single pair of properties that divide reasoning into two natural kinds, and so their views are monothetic. These accounts differ in two crucial respects. First, Carruthers's assumes that reflective reasoning and intuitive reasoning (themselves distinct kinds of processes) are subserved by reflective and intuitive systems respectively. Stanovich claims that there are two systems (the algorithmic mind and reflective mind) that carry out Type-2 processing and many systems that carry out Type-1 processing, while Evans wishes to remain agnostic as to how many systems carry out Type-2 processing. Second, the way in which these theorists divide the two

[12] One might object: if the new division, which dual-process theorists use to displace the Standard Menu, leaves out the cross-cutting properties in favor of some reduced set of properties, this is no problem. However, what is doing the work in this reply is the limiting of properties rather than the switch from *system* talk to *process* talk.

[13] Evans and Stanovich's definitions of autonomy are very similar to the way that many would define 'automatic.' Since Evans and Stanovich talk about 'autonomous processes' instead of 'automatic processes,' I will do the same when addressing their accounts.

kinds are incompatible with one another, given the abandonment of the Standard Menu: Carruthers claims that the real distinction is between intuitive and reflective processes whereas Evans and Stanovich claim that the real distinction is between autonomy and those involving working memory. Carruthers says that intuitive processes may "nevertheless be employing working memory to process the task instructions and maintain the ensuing representations long enough for the intuitive systems to generate an answer" (2013a, p. 20). (see chapter 1 for more detail on these accounts).

One of the virtues of the Standard Menu was that it unified the dual-process theories. In his reply to Evans and Stanovich (2013a), Keren (2013) says he is reminded of a Russian folktale in which a fool is taught to make "stone soup" by boiling a stone in water. This alone is sufficient for making that water into stone soup, though one could add any kind of meat or vegetables in order to improve taste. Keren says that "inspecting the different labels proposed and the various terminologies employed to characterize the presumed two systems and their corresponding alleged processes strongly suggest that it has become a stone soup where everything goes" (p. 257). Dual-process theorists might once have replied by citing their shared allegiance of the general way that they divided the two processes. There might have been minor disagreements as to which properties should be cut from the Standard Menu, but these theorists could always point to the many properties of the Standard Menu which they held in common with one another. Now that this recourse is undercut, and if these theorists continue to call themselves "dual-process theorists," the term indeed begins to look like a stone soup. [14]

3.3. What is a reasoning process?

The broader that we cast 'reasoning,' the more plausible it is that reasoning processes are of more than one kind, but the less interesting the claim that reasoning is of two or more kinds becomes.[15] Thus, while it is difficult for anyone to define reasoning, it is a pressing issue for dual-process theorists. Because there are many properties on the Standard Menu, many processes

[14] It is not the critics who insist on maintaining the 'dual-process' label. Dual-process theorists continue to use it. Furthermore, dual-process theorists seem to regard each other as allies, and, at times, downplay the differences between their own versions of dual-process theory.

[15] Furthermore, note that the broader we cast "reasoning," the more likely it will be that there are more than just two kinds of processes.

were excluded from the S1/S2 distinction because they did not fit into either category. In a way, then, the Standard Menu drew boundaries around the concept of reasoning and also divided reasoning processes into two kinds. However, Evans, Stanovich, and Carruthers's accounts fail to draw boundaries around reasoning, which threatens to trivialize their accounts.

Let me begin with Evans and Stanovich. The instances of belief formation paradigmatic of Type-1 processes are indeed autonomous in Evans and Stanovich's sense. However, many autonomous processes are not reasoning processes at all. As Sloman (2014) notes, Stanovich & Evans "are casting their net too wide. The vast majority of what goes on in the body and the brain meet this definition of Type-1 processing including (say) laughing when being tickled" (p. 71). Although this is merely a passing comment, Sloman reveals an important way in which Evans and Stanovich's new account is weaker than the Standard View. Let me say why: absent-mindedly driving a car, sneezing, and breathing are all autonomous, but are not reasoning processes. As such, absent-minded driving, sneezing, and breathing should not count as Type-1 processes since Type-1 processing was supposed to be about *reasoning.* If one uses the Standard Menu to characterize Type-1 processes, then these kinds of processes are ruled out, but on Evans and Stanovich's monothetic accounts, they are not.

Carruthers's account faces a similar problem: unconscious processes surely include the vast majority of what goes on in the mind. Again, reflexes, absent-minded driving, and breathing are all unconscious. However, these unconscious processes should not count as intuitive reasoning processes.

There are two replies, both of which will require some alternative way to distinguish reasoning and non-reasoning processes. First, these theorists might admit that Type-1 or intuitive processes are indeed pervasive: absent-minded driving, sneezing, and breathing are Type-1 processes/intuitive processes. Problematically, if the concepts Type-1 and Type-2 or intuitive and reflective apply so broadly, then it becomes trivial that there are Type-1 and Type-2 processes or intuitive and reflective processes. It is true that my breathing is a different kind of process than my construction of a counterfactual possibility. How could any of us doubt this? Problematically, this watered down version of dual-process theory seems compatible with several one-system accounts of reasoning such as Osman (2004) or Kruglanski and Gigerenzer (2011) (in the next chapter, I will briefly examine these accounts).What made dual-process

theory so interesting was the radical claim that reasoning itself is divided into two kinds of processes and (on some accounts) underwritten by two very different kinds of cognitive systems.

Perhaps the dual-process theorist will reply that the interesting dual-process claim is that some Type-1 processes are reasoning processes. In other words, that there are some reasoning processes that are autonomous (or unconscious) in the same way that absent-minded driving, sneezing, and breathing are autonomous (or unconscious). However, to assess whether this claim is true, we would need some principled way, which we currently do not have, of determining whether or not a process is a *reasoning* process.

Second, dual-process theorists might claim that Type-1 or intuitive processes are only meant to mark a distinction *within* reasoning. That is, Type-1 processes are *reasoning* processes that are autonomous, or intuitive processes that are unconscious *reasoning* processes. This reply avoids the above objection, since most autonomous or unconscious processes are not *reasoning* processes, but in order to assess the truth or substantiveness of this claim, we need to know the boundaries of the concept 'reasoning' such that reasoning is supposed to be divided into two neat kinds. Again, the broader the extension of the concept reasoning, the more plausible it is that reasoning is divided into more than one kind (perhaps more than two kinds). However, the broader the extension of the concept 'reasoning,' the less interesting becomes the claim that reasoning is of two or more kinds.

3.4. A dilemma: loss of the promise of explanatory power or falling back on the Standard Menu

I will argue that, in moving to monothetic accounts, Stanovich, Evans, and Carruthers's new theories lack the promise of explanatory power that the Standard View possessed. Evans, Stanovich, and Carruthers face a dilemma: either their account lacks the explanatory power that the Standard View promised (and, thus, inferences to the best explanation for dual-process theory are undercut), or they must tacitly assume properties on the Standard Menu cluster when offering explanations. In practice, Carruthers has taken the former horn of this dilemma, while both Evans and Stanovich have fallen into the latter.

Let me begin with the first horn: Evans, Stanovich, and Carruthers's accounts lack the explanatory power promised by the Standard View. On Evans and Stanovich's accounts, autonomy does little explanatory work on its own. Remember that Evans & Stanovich's (2013a)

definition of autonomy follows Fodor's (1983) definition of automaticity: autonomous processes are "mandatory when their triggering stimuli are encountered and they are not dependent on input from high-level control systems" (2013a, p. 236). Now consider experiments that have been taken to support dual-process theory because of the plausible explanation that dual-process theory offers. However, let us only use the mandatory/controlled distinction. Consider the classic example of the representativeness heuristics: Linda the bank-teller (Tversky and Kahneman 1983). What does the Type-1 processing explain here? At best, it explains why it is that one response "beats out" a second response. One response (which happens to be the incorrect one) "comes to mind" more quickly (since it is mandatory) than a controlled process (which requires the use of working memory). Since Evans and Stanovich's accounts are default-interventionist, and since subjects are cognitive misers, it may be that no Type-2 process is initiated. But this does not offer an explanation for why most subjects say that Linda is more like a feminist bank-teller—it only offers an explanation for why most subjects offer a specific response: a Type-1 process is mandatory, and so the Type-1 response will "come to mind" regardless of what the subject does. However, this explanation fails to answer why it is that subjects respond in the way that they do: why is it that subjects tend to say that Linda is more likely a feminist bank-teller?

Crucially, the lack of an explanation persists even when we include those properties that Stanovich and Evans (2013a) say will be closely correlated with autonomy: that the process is *fast, efficient,* and *parallel* does not help explain why subjects tend to deliver the response that they do. The problem is not merely that the monothetic properties (i.e. autonomous/working memory involving) alone fail to provide an explanation for results often taken to support dual-process theory: even those properties that are supposed to correlate with autonomy cannot do the explanatory work needed to motivate dual-process theory. Since the natural kinds posited by dual-process theory were introduced to explain why subjects tend to deliver the responses they do in experiments like the Linda case, the very reason for positing Type-1 and Type-2 processes as natural kinds has been undercut.

The obvious reply for the dual-process theorist is to say that autonomous processes are heuristic or associative. The description of Linda "fits better" with the claim that she is a feminist than that she is a bank-teller. This might be because a feminist is associated with words in the description of Linda, or it might be that the Type-1 process utilizes a representativeness heuristic.

This might work as an explanation, but only by using properties from the Standard Menu that are supposed to cross-cut the autonomous/working memory involving distinction. In practice, this is exactly what Stanovich and Evans do. While they tell us that Type-1 and Type-2 processes are distinguished using a singular property pair, they then assume that there is a clustering of properties along the lines of the Standard Menu in their explanations. Importantly, the properties needed to do the explanatory work (such as associative or heuristic) are exactly the properties they claim do not cluster with the new distinction they draw. This is the second horn of the dilemma.

After telling us that being autonomous is a necessary and sufficient condition for being a Type-1 process, Stanovich (2011) goes on to claim that autonomous processes include: "behavioral regulation by the emotions; the encapsulated modules for solving specific adaptive problems that have been posited by evolutionary psychologists; processes of implicit learning; and the automatic firing of overlearned associations" (p. 19-20). Stanovich defines Type-2 processing using the contrary of each property he used to define Type-1 processing. Thus, Type-2 processing is non-autonomous, slow, does put pressure on central computing, is serial (i.e. not parallel), and is often language-based (2011, p. 20). All hypothetical thinking is Type-2 processing, though the converse does not hold (2011, p. 47). This way of identifying Type-1 processes begins to look like Stanovich's (1999) cluster proposal, since Type-1 processes are autonomous, modular (and so evolutionarily old and fast), and heuristic. So at least some properties from the Standard Menu remain in the account, and, as such, will be open to the problematic cross-cutting cases that made Evans and Stanovich move away from Type-1/Type-2 processing as cluster kinds. Again, recall that Evans (2006, 2008) and Carruthers (2013a) have explicitly argued that autonomous processes need not be associative or heuristic processes. Perhaps Stanovich is an outlier here, wanting to maintain that associative or heuristic processes do correlate with autonomous processes. However, if Stanovich wishes to maintain that associative, heuristic, and autonomous processes cluster, then it is incumbent on him to argue against Evans and Carruthers who have provided evidence to the contrary.

As with Evans and Stanovich's suggested distinction between autonomous and controlled processes, Carruthers's distinction lacks the promise of explanatory power and experimental evidence that was supposed to make dual-process theory so attractive. First, since Carruthers's

distinction between intuitive and reflective processes amounts to the difference between unconscious and conscious processes, the intuitive/reflective distinction does not add any explanatory power for those of us who already thought that there exists unconscious and conscious processes. Worse still, this new distinction lacks the power to account for the explanandum of dual-process theory. Suppose we know that a process is conscious; this alone does not tell us much about the resulting output of that process. It is hard to say exactly what *functional* difference consciousness makes to the output of a process.[16] Surely consciousness is not epiphenomenal, but it would be odd to claim that *consciousness* is the difference-maker between subjects' varying responses in the reasoning and decision-making literature. In fact, since Carruthers argues that reflective processes can employ heuristics and the performance of intuitive processes sometimes approximates "that of an ideal Bayesian reasoner" (2013a, p. 6), it is clear that the intuitive/reflective distinction cannot explain why subjects tend to respond incorrectly in so many of the paradigmatic experiments from the reasoning and decision-making literature. Thus, it is unclear, for two reasons, what explanatory power Carruthers's account buys us. First, it amounts to the conscious/unconscious distinction, and so adding intuitive/reflective to our mental ontology does not add explanatory power. Second, the conscious/unconscious distinction cannot explain the data from the heuristics and biases literature.

One might attempt to find a way out of the dilemma as follows: *autonomy* and *working memory involving* are natural cognitive kinds, each of which corresponds to a cluster of properties, and these categories are therefore projectible and explanatorily powerful. Since a process is Type-1 if and only if it is autonomous, Type-1 processing is a natural cognitive kind as well, and, likewise, since a process is Type-2 if and only if it involves working memory, Type-2 processing is also a natural kind. Carruthers might reason similarly, *mutatis mutandis*: *unconsciousness* and *consciousness* are cognitive kinds, so *intuitive* and *reflective* processing are cognitive kinds as well. For the sake of argument, I will assume that autonomy, working memory involving, conscious, and unconscious are natural kinds. Responding to this line of argument requires that I treat Evans and Stanovich separately from Carruthers. I begin with Evans and Stanovich.

[16] This difficulty remains even if we reject the possibility of zombies or that Mary learns something new.

In order for the above inference (from *autonomy* and *working memory involving* being natural kinds to Type-1 and Type-2 processes being natural kinds) to be valid, it needs to be the case that Type-1 processes are *identical* to autonomous processes, and Type-2 processes must be *identical* to working memory involving processes. To see why, consider the above argument: a process is Type-1 if and only if that process is autonomous. Autonomy is a natural kind. Therefore, Type-1 is a natural kind. The argument is invalid unless we strengthen the first premise to: Type-1 processing is autonomous processing.[17] However, in order to have a substantive empirical identity claim (e.g. water=H_2O) we must have some understanding of each half of the identity claim. However, if Evans and Stanovich defend their view by *identifying* Type-1 and Type-2 processes with *autonomous* and *working memory involving* processes respectively, then we do not have an independent understanding of each half of the identity claim, since we have no handle on the extension of the concepts Type-1 and Type-2 apart from the stipulations of dual-process literature. We cannot introduce a new natural kind (X) into our ontology merely by saying that X is identical to some known natural kind.

Perhaps Evans and Stanovich might say that they are fine with all of this. Type-1 and Type-2 processes are not *new* natural kinds. Rather, the substantive claim is that the Type-1/Type-2 (i.e. autonomous/working memory involving) distinction is superior to the S1/S2 distinction, and, in displacing the S1/S2 distinction, dual-process theory has made progress. That is, Type-1 and Type-2 are successor concepts for S1 and S2. However, now the dual-process theorist runs back into the first horn of the dilemma: their successor concepts promise less explanatory power than the S1/S2 distinction promised.

Even if Carruthers succeeds in establishing natural kinds, he does not succeed in establishing successor concepts for S1 and S2. Again, in order for the inference (from *unconscious* and *conscious processes* being natural kinds to *intuitive* and *reflective processes* being natural kinds) to be valid, it needs to be the case that intuitive processes are *identical* to unconscious processes, and reflective processes must be *identical* to conscious processes. Carruthers's view does not run into my first objection, since we have some pre-theoretical

[17] To see further why Type-1 and autonomous processes would be identical, consider that cluster kinds are identified by a cluster of properties. Thus, if Type-1 and autonomous are perfectly correlated (as Evans and Stanovich claim), then Type-1 could just as easily be identified by the cluster of properties that identifies autonomy. Since the same cluster would identify Type-1 and autonomous processes, Type-1 and autonomous processes would be identical.

understanding of intuitive, reflective, unconscious, and conscious processing. Now, it seems that Carruthers intends his distinction as a successor concept, since he takes the Standard Menu as his point of departure, and he himself adopted the S1/S2 distinction in the past (see Carruthers 2009). However, for his distinction to be a successor to the S1/S2 distinction, it should provide similar explanatory promise as the S1/S2 distinction. This drives Carruthers back towards the first horn of the dilemma: the intuitive/reflective distinction lacks the same explanatory promise as the S1/S2 distinction. If Carruthers does not intend intuitive and reflective as successor concepts, then it is unclear why the intuitive/reflective distinction would need to be introduced at all, since the kinds to which intuitive/reflective are identical (i.e. the kinds unconscious and conscious respectively) are already in our ontology. I conclude that using the clusters of properties that correspond to the (putative) natural kinds autonomous, working memory involving, unconscious, or reflective will not help these accounts.

4. Conclusion

Even if the distinctness claim is true, it is unclear how we would divide the two kinds of reasoning. I have argued that there is reason to be skeptical of the cluster kind claim, since the properties on the Standard Menu crosscut each other. In response to this problem, Stanovich, Evans, and Carruthers have adopted a monothetic way of distinguishing kinds. However, the recent move by dual-process theorists to monothetic accounts will not help. First, in rejecting that the properties on the Standard Menu cluster, Evans, Stanovich, and Carruthers lose a way to limit what we conceive of as reasoning. These theorists might reply that their distinctions were never meant to define 'reasoning processes.' However, these theorists owe us a way to draw boundaries around reasoning in order to avoid threats of triviality. More importantly, these new ways of dividing cognitive processes lack the explanatory promise of the Standard View that made dual-process theory attractive as an account of natural cognitive kinds. As a result, both Evans and Stanovich sometimes fall back into using properties from the Standard Menu that they claimed did not correlate with their new distinction. They thereby tacitly commit themselves to the Standard View. Assuming that the Standard Menu cannot distinguish two kinds of reasoning, then, we would be wise to abandon dual-process theory.

Chapter 5: Toward a One-System Account of Reasoning

In the last three chapters I have been arguing against, or undermining the arguments for, two-system and dual-process theory. In chapter 2, I offered the rough outline of a one-system account of reasoning and argued that it can explain the data from the heuristics and biases literature just as well as two-system theory. In chapter 3, I argued that the distinctness claim is empirically testable in that the one-system theory is incompatible with SCB arising from a single simultaneous reasoning processes, and I offered a way to test for the existence of such SCB. Apart from parsimony, I have offered no positive argument against the distinctness claim; I have only tried to undercut arguments for the distinctness claim, and the distinctness claim remains epistemically possible (this is, after all, an empirical question). However, given my argument in chapter 4 that the properties used to distinguish S1 from S2 (as well as some ways of distinguishing Type-1 processing from Type-2 processing) cross-cut one another, even if the distinctness claim is true, the systems (however many there are) will not be divided as two-system theorists have maintained.

It is time to fill in the details of my one-system alternative to two-system and dual-process theory. I begin by offering a necessary condition on 'reasoning.' I then outline which of the properties from the Standard Menu my one-system account retains. In order to assuage the worry that the explanations my theory offers of the heuristics and biases literature are *ad hoc*, I will outline a taxonomy of reasoning errors. It is crucial to my account that I explain mode determination of the reasoning system. In short, I argue that the modes are determined by a combination of environmental factors (including wording of the problem, setting, and practical considerations such as time) and individual thinking dispositions. I argue that most mode determination occurs unconsciously. I further examine how my one-system theory might accommodate cognitive decoupling in reasoning, which is how the reasoning system moves from concrete reasoning to abstract reasoning. My taxonomy of reasoning errors and explanation of cognitive decoupling are meant as rivals to Stanovich's appropriation of cognitive decoupling to his tripartite division of reasoning and his dual-process theory laden taxonomy of human reasoning errors. Finally, I empirically distinguish my own account from Evans and Stanovich's default-interventionist accounts and offer evidence that favors my theory over theirs.

1. Getting clear on 'reasoning'

Two-system theorists have drastically expanded the concept 'reasoning.' Of course, the broader the concept of 'reasoning' applies, the less plausible that only one system is responsible for all reasoning processes. However, as Samuels (2009) points out, if we adopt a wide understanding of reasoning, it is implausible that only *two* systems (or processes) can accommodate all the tasks which fall under the broad concept of 'reasoning.' I begin this section with Samuel's attempt to define reasoning in a way amenable to two-system theory. We need, at the very least, a necessary condition of reasoning to exclude certain kinds of processes. To this end, I defend Samuels's suggestion that *inference taking* is a necessary condition on reasoning.

Samuels (2009) attempts to define reasoning such that it is plausible that there are exactly two systems, and argues that there is no such characterization of reasoning. While his purpose is to argue that the theorists might maintain the S1 and S2 distinction only in a type form, a further upshot of his argument is that we do not, at present, have any way of defining 'reasoning,' and, as such, we may have to rely on our intuitive notion of reasoning. He offers five ways of defining reasoning.

Suggestion 1: All systems that draw inferences are reasoning systems. This includes too many systems, since inference-making is pervasive in cognition (and indeed, in perception). Samuels claims that it is a plausible necessary condition on reasoning, and I will argue below that he is right. One might attempt to exclude non-reasoning systems which engage in inference making by being more specific about what kind of inference is required, which leads us to Samuels's next suggestion.

Suggestion 2: "Any system that subserves conscious deliberative inference is a reasoning system" (Samuels 2009, p. 135). This definition excludes too much. In particular, it excludes S1 from being a reasoning system (or, at least, it excludes certain members of TASS from being reasoning systems).

Suggestion 3: "Any device involved in paradigmatic reasoning tasks is a reasoning system" (Samuels 2009, p. 135). Again, this includes too much. Consider all the task that 'involve' reasoning: "arithmetic inference (Dhaene 1997; Gelman and Butterworth 2005), probabilistic reasoning, decision making, planning, spatial reasoning, reasoning about social phenomena, ethical judgment (Greene and Haidt 2002), reasoning about the minds of others

(Leslie et al. 2004), enumerative induction, and abductee inference" (Samuels 2009, p. 136). Furthermore, there are language systems, perception, and motor control 'involved' in some of these tasks. If being 'involved' in a reasoning process is sufficient for being a reasoning system, then language systems, perceptual systems, and motor control systems will count as reasoning systems.

Suggestion 4: "Reasoning systems are to be identified with so-called 'central systems'" (Samuels 2009, p. 135). First, it is hard to get clear about what these central systems are. Samuels (1998) has argued that the central systems are those that subserve reasoning, but even if Samuels is right, it will not help here, since reasoning is precisely what were are attempting to characterize.

Suggestion 5: We may rely on our intuitive understanding of 'reasoning' without defining it. Samuels suggests that this might be the right way to go, but it would be unlikely that only two systems are involved in all the tasks which we would 'intuitively' call reasoning. Consider the long list quoted in response to suggestion 3 to see why.

Perhaps one reason that it is so difficult for dual-process theorists to find a definition of reasoning that will work of their theory is precisely because they have been too liberal in their application of the term. For example, they seem willing to let linguistic understanding count as a Type-1 process/a process carried out by S1. Linguistic understanding is fast and automatic. Two-system theorists have taken priming (both with numbers and words) to support their theories, and priming is supposed to arise from S1/Type-1 responses. Since S1 is a reasoning system, its processes will be reasoning processes. However, the process whereby sounds or writing are interpreted as meaningful is not a type of *reasoning*.

Logicians might be reluctant to call most of what S1 engages in 'reasoning.' One might go so far as to say that 'reasoning' just is engaging in deductive and inductive inferences. Notice that the process that results in responding to a logic problem is not necessarily a reasoning process. Consider that I might respond to the Linda case in a way that does not involve reasoning. For example, I could flip a coin, which I know has no bearing on Linda's being a bank-teller or feminist. A good candidate for ruling out this kind of case is to say that it fails to be a reasoning process because I would not have formed my response by way of *inference*. Thus,

it is a necessary condition on being a reasoning process that subjects form their response by way of inference. Call this the **inference criterion**.[1]

One might provide the following counterexample to the inference criterion: A subject might reply to the Linda problem by way of a coin toss. However, while there must be some deliberative process leading me to coin flipping, the subject need not make an *inference.* After all, sometimes we do make choices based on the flip of a coin: in the case where it makes no difference to me whether I choose A or B, I might let a coin decide. According to this objection, the coin flip is a reasoning process, but does not involve inference. Perhaps the following is a reasoning process: deciding to flip a coin because neither option is better, flipping the coin, and acting accordingly.[2] While these collective events constitute a reasoning process, this does not imply that each part of the process is a reasoning process. This whole process would be a reasoning process in virtue of *deciding* that a coin flip is appropriate, but that decision would involve inference. The subject might reason as follows: I do not care which of the two options I choose (or I might want the process to be random), I recognize that a coin toss would accomplish just this, and so I infer that flipping a coin would be a good method. So the whole process does involve an inference. On an alternative interpretation, one might be reluctant to call the whole process (deciding to flip, flipping, and responding) a reasoning process at all. One might think that only a subset of theses constitutive events is a reasoning process. In that case, part of a process might involve reasoning while other parts of the same process do not. Thus, whether we regard the whole process (deciding to flip, flipping, and responding) as a reasoning process or not, it is no threat to my inference criterion, since if we do so it would be because some aspects of the whole process involve inference, namely the decision to flip a coin.

Notice that I have narrowed the concept of reasoning, and so the explanandum for which my account is narrowed as well. I am not changing the subject; I am elucidating the concept. If we restrict the concept 'reasoning' according to the inference criterion, then the two-system theory may lose some of its appeal. I propose that in building our cognitive architecture of

[1] Although I leave 'inference' unanalyzed here, the inference criterion is still informative because it rules out some processes as non-reasoning processes.

[2] Recall that I take a process to be a collection of events that are causally connected in a certain way (as opposed to being grouped as a mere matter of convention). See chapter 3.

human reasoning we restrict ourselves to processes of inference. Having clarified what the nature and extension of 'reasoning,' I now turn to how reasoning operates on my one-system theory.

2. Modes of operation on the one-system theory

Reasoning must involve inferences. What kind of ways might the reasoning system draw inferences? In chapter 2 I proposed that the architecture of human reasoning is of a single reasoning system which operates in different modes. Again, I take modes to be properties (see Heil 2003, Martin 2008). Most generally, modes are ways that things are. Here I am concerned primarily with functional properties.[3] The reasoning system operates in various ways, and each of these ways is a mode. Thus, modes of the reasoning system are ways that the reasoning system operates.[4]

In chapter 2, I claimed that the reasoning system can operate consciously or unconsciously, automatically or in a controlled manner (though this is a matter of degree), and inductively or deductively. Furthermore, as I argued in chapter 4, these properties do not cluster. For example, the reasoning system might frequently solve a problem through its deductive mode combined with its automatic mode. Thus, in addition to being metaphysically different than the two-system theory in that it posits only one entity whereas the two-system accounts posit two, my account allows for the properties on the Standard Menu to cross-cut each other. I am not claiming that there is one reasoning system that operates in a Type-1 fashion sometimes and a Type-2 fashion at other times (i.e. denying the Distinctness Claim while endorsing the Kind Claim applied to processes). I am denying both the Distinctness and Kind Claims.

For any reasoning process and any property pair, the reasoning system will operate in a definite mode. For example, it will not operate associatively and in a rule-based manner at the same time. As a way of illustration, we might imagine our reasoning system as an audio mixing

[3] Or the functional aspect of properties, if you think, as I do (Mugg 2013), that functional and qualitative properties are not ontologically distinct.

[4] Stanovich and Evans (2013a) claim that modes and properties are distinct. Properties are differences in kind while modes are differences in degree. Properties distinguish kinds of processes, such as Type-1 and Type-2, but there are several modes of operation within Type-2 processing. The reason that Type-2 processing does not break down into further kinds is because opposing modes are of degree rather than kind. Oddly, some of those properties used to distinguish Type-1 from Type-2 are differences in degree rather than kind (fast/slow and (I have argued) automatic/controlled). However, the differences in terminology between myself and Stanovich and Evans should not be one of concern.

board with several switches and slides: one for each property pair. What sound the mixing board outputs for a given sound input depends upon where each of the slides are positioned and which way each switch is flipped. In the same way, the reasoning system's output depends upon which modes it is in. Furthermore, one does not simply set each of the slides and switches on a mixing board and then run sound through it. Rather, as one listens to the output sound, one manipulates the slides and switches until the output sound is as desired. This is to say that there is feedback. In a similar way, there is feedback in the case of the reasoning system. While some details of the reasoning problem might put the reasoning system into certain modes (e.g. the word 'probably' puts it into an inductive mode (see chapter 2)), the reasoning system can change modes as it is working through a problem. There is not necessarily a single route that the reasoning system is forced into every time, rather the system is dynamic.[5] For example, if the stakes are high and the subject has time to check his or her answers, the reasoning system may go back over the answer in a rule-based way that demands more cognitive resources.

Which of the properties from the Standard Menu should we retain in the one-system theory? Exactly which properties from the Standard Menu we must abandon is a matter for future research, but there are some properties that should certainly remain and others that need to be removed. Some of the property pairs on the Standard Menu are on a continuum (such as automatic/controlled, fast/slow, conscious/unconscious), and some are not (inductive/deductive, associative/rule-based). In the last chapter, I surveyed the empirical literature to clarify some of the key concepts from the Standard Menu, namely associative, rule-based, automatic, controlled, fast, slow, evolutionarily old, and evolutionarily new. I do not retain all the properties on the Standard Menu, or even these properties, in my one-system account. For example, the evolutionarily old/evolutionarily new distinction is, I think, unhelpful. A process or system being evolutionarily old or new is not intrinsic to the process or system itself; it is merely an indication of the origin of the system in question. While the way in which a reasoning system evolved is an interesting question on its own, it does not tell us how the reasoning system operates, even if the two issues are related. Evolutionary age can only be an indication of what kind of functional

[5] Although some have used dynamic theories of mind to argue against the need for representations (see Chemero 2000, Brooks 1991, 1999, and van Gelder and Port 1995), all I mean to imply is that there is feedback.

properties are at work, and we would likely determine evolutionary age by looking at those functional properties as they appear in the living phylum.[6]

Some properties on the Standard Menu are higher-order in the sense that a system that operates in that way depends on that system operating in certain other ways. To put it a little more precisely: M mode is higher-order just in case, both 1) if a system S operates in M then S operates in modes M_1-M_n, and 2) no single mode M_1- M_n is individually sufficient for M.[7] For example, whether a system is evolutionarily rational or epistemically rational depends upon its functional characteristics, rather than certain functional properties depending (ontologically) on a system being evolutionarily rational or epistemically rational. Being in certain lower-order modes *constitutes* being in certain higher-order modes. For example, being in a mode that requires the use of executive functioning and working memory constitutes being in a high-effort mode (as opposed to a low-effort mode), and operating on the basis of implicit associations constitutes operating stereotypically. While dual-process theorists who advocate the Cluster Kind Claim can easily add these higher-order properties to the Standard Menu, I cannot. Dual-process theorists can add higher-order properties to the Standard Menu because the properties upon which the higher-order properties depend are supposed to cluster. For example, if 'being an evolutionarily rational process depends upon a system being fast, associative, and automatic, then S1 will be evolutionarily rational because these properties fast, associative, and automatic cluster. Thus, on two-system accounts of reasoning, higher-order properties can do much explanatory work, since each system will possess several higher-order properties.

If the reasoning system frequently operates in the higher-order modes, then the lower-order properties cluster. However, I have argued that these lower order properties do not cluster. Thus, we should not expect that the reasoning system will frequently operate in these higher-order modes. To use the audio mixing-board analogy again, it is unlikely that my one reasoning

---

[6] One might think that evolutionary age can illuminate cognitive architecture by showing which capacities are not necessarily connected. I disagree, since it is not *evolutionary age* that does the heavy lifting in such arguments. Suppose that someone claims that capacity A requires capacity B, but then we find that there are species that possess capacity B but not capacity A. We can conclude that A does not require capacity B. While the argument is perfectly valid, *evolutionary age* did not figure in it. What is important is the determination that the capacities are dissociable as demonstrated in the living phylum.

[7] The concept of dependence is a thorny one, and perhaps cashing it out using material conditionals is problematic. The above formulation should be sufficient for my purposes, but it will make no difference to my argument if you substitute a more baroque account of dependence.

system will have a ‘switch’ for these higher-order properties. Of course, the reasoning system might operate in a higher-order way, but it will be the lower-order properties—upon which that higher-order property depends—that the one-system switches in and out of. Thus, talk of these higher-order properties are only useful shorthand for the lower-order properties. As such, I will only focus on the lower-order properties.

Some of the binary distinctions I retain from the Standard Menu are: unconscious/conscious, concrete/abstract, and mandatory (in the sense that once it is started it cannot be stopped)/controlled. My one-system theory will further rely on the inductive/deductive distinction which the S1/S2 distinction cross-cuts. Other properties on the Standard Menu (such as automatic) are less helpful than the sub-properties composing those properties. Following Kunda (1999), I break ‘automatic’ into its conceptual components. Automatic processes are mandatory, unintentional, possibly parallel, highly efficient (which admits degrees), and outside awareness. Processes can be more or less automatic because a process might only possess some of these properties. Instead of using the ambiguous concept ‘automatic,’ I will use these concepts individually, since the concepts used to define automaticity and their antonyms should already be clear.

Some of these binary distinctions from the Standard Menu might break down further. For example, abstract reasoning itself can be of a few kinds. First, it might be algorithmic. A logic student who mechanically translates a syllogism into sentential logic and then recognizes an instance of modus tollens is operating abstractly and algorithmically. On the other hand, a student who has only been taught sentential logic and is given a syllogism with quantifiers (‘all’ and ‘some’) might attempt to construct Venn diagrams to judge whether the argument is valid. This would be a case of abstract reasoning that is not algorithmic.

The reasoning system may demand the use of executive functioning; it may use cognitive resources. I do not include a distinction between ‘using cognitive resources/not using cognitive resources’ (which is Evans and Stanovich’s (2013a) way of distinguishing Type-1 and Type-2 processes) because the amount of cognitive resources a reasoning process used depends upon the modes in which the reasoning system operates, rather than being a mode of the system itself.

An important question for my one-system theory is when and how the reasoning system changes from mode to mode, especially through feedback. This remaining question is analogous

to the question for two-system theory as to how and when an S2 response overrides a S1 response. The Frame Problem (as interpreted by Dennett and Fodor) is that of explaining how a creature recognizes which beliefs he or she ought to re-evaluate after that creature has engaged in some course of action (see Dennett 1978, p. 125 and Fodor 1983, p. 114). It should be unsurprising that a cognitive architecture for reasoning runs into the Frame Problem, since it is "as deep as the analysis of rationality" (Fodor 1987, p.140). Analogously, the current (related) problem is that we must account for how the reasoning system determines what mode it ought to be in in order to deliver the appropriate response for a given reasoning problem. The problem facing my one-system account is that of a regress: the reasoning system must operate in some mode to solve a reasoning problem. However, deciding which mode is best for solving the problem is itself a reasoning problem. Thus, if the reasoning system itself is responsible for choosing the modes it uses to solve reasoning problems, then we have a regress.

We may block the regress by denying that mode determination is a matter of actively choosing. Indeed, I think we must deny that the reasoning system alone determines which of the initial modes it will use in solving a problem. Instead, my suggestion that the reasoning system is usually set in certain modes by environmental factors (i.e. setting of test, wording of problem, etc.), rather than actively altering itself prior to entering into the reasoning process. Thus, the reasoning system does not go through a mode determination stage, which would be a kind of reasoning process, and no regress occurs. However, environmental factors cannot be the sole determining factor: individual thinking dispositions must be a consideration as well (these include how much working memory capacity individuals possess and how much subjects are interested in working on specific kinds of problems). A complete account of mode determination on the one-system theory will include a full picture of which environmental and individual thinking dispositions alter mode determination. I make no claim to have offered a comprehensive account. Future research should focus on identifying what it is in the environment that puts the reasoning system into which modes and when and how a subject make's a conscious choice as to how to reason. I have made some suggestions throughout this dissertation (especially in chapters 2 and 3). While the data from the heuristics and biases literature has given us data on mode determination, we need to be more careful to control which modes are being altered.

While many subjects respond incorrectly in classic experiments from the heuristics and biases literature, it is important to note that some respond correctly, and any account of human reasoning must account for this, which Stanovich calls 'individual difference.' Interestingly, the existence of individual difference has been an objection to some two-system theories, according to which it is unclear why some subjects, but not all, subjects are able to override their S1 response. If my one-system alternative is to have any force, it must provide a response as well. Stanovich further provides us with a framework for conceptualizing individual reasoning differences on his tripartite division of the mind (see below). My one-system alternative should be able to provide something similar.

One reason (but not the only reason) that humans err in reasoning is that the reasoning system sometimes operates in a suboptimal (though not necessarily 'incorrect') mode. It may (for example) operate inductively when it could more optimally operate deductively. It may operate under its mandatory mode unconsciously, and thus the subject might not check his or her answer. Notice that unconscious or inductive reasoning may very well lead to the correct response even if they are suboptimal in certain cases. So my suggestion is not tautological. One way in which subjects become better reasoners is through gaining the ability to put their reasoning system into the proper mode for each task. One of the most important modes to switch into is the abstract mode, especially in novel circumstances and when the subject has a liberal amount of time to produce a response. That is, for generalizable knowledge, subjects who can abstract away from the particulars will be able to reason more in line with full rationality than those who are unable to abstract away from the particulars. A necessary condition on switching into abstract reasoning is cognitive decoupling—which Stanovich (2011, 2012), Evans and Stanovich (2013a), Frankish (2010, 2012), and Carruthers (2006, 2009, 2013a, 2013c) all emphasize in their theories. I will focus on Stanovich's (2011) influential account. Cognitive decoupling occurs when subjects abstract away from the details of a problem, especially their own subjective perspective of the problem. Cognitive decoupling is important for one-system and two-system account of reasoning because it is through cognitive decoupling that one moves from concrete reasoning to abstract reasoning. Cognitive decoupling allows subject to attend to the form of an argument rather than the content. Thus, it is essential in properly determining validity of an argument. On a two-system account, cognitive decoupling will involve moving either from an S1/Type-1 process to a

S2/Type-2 process, or else a move into the 'reflective mind.' Of course, the mechanisms of cognitive decoupling will differ between the one and two-system accounts. For example, on my account, cognitive decoupling will not be identified with only one set of properties. Before explaining how cognitive decoupling works on my one-system account, and how it figures into a taxonomy of reasoning errors, it will be helpful to outline how Stanovich accounts for cognitive decoupling within his tripartite account of reasoning.

3. Cognitive decoupling

Cognitive decoupling is a key mechanism by which a subject moves from concrete reasoning to abstract reasoning. It is necessary for attending to relevant details of a problem—such as the structure of the argument—while ignoring irrelevant details of the problem. Cognitive decoupling allows subjects to avoid several biases: myside bias, belief bias, and curse of knowledge effect, to name a few. On Stanovich's account it *is* the way in which subjects move from algorithmic thinking to reflective thinking. Cognitive decoupling is the process whereby subjects abstract from real world particulars. This concept has been deployed in various domains, and I will focus on its use in both philosophy of mind (Nichols and Stich 2003, Dennett 1984, Carruthers 2000, 2006) and developmental psychology (Leslie 1987, Perner 1991). I begin with the latter. Leslie (1987) uses cognitive decoupling to explain the emergence of pretend play in children (e.g. this whiffle-ball bat is a sword). According to Perner, there is a primary representation of the world and a copy of the primary representation (and so it is a secondary representation). Children use the secondary representation to engage in pretend play. It is because the secondary representation is not generated by the world, but is generated from a *representation of the world*, that children are able to treat the second representation[8] as pretend. Leslie claims that the secondary representation is needed to avoid becoming confused as to what is real and what is pretend. He also points out that the existence and use of the secondary representation leaves intact the primary representation, and so the primary representation "is free to continue exerting whatever influence it would have on ongoing processes" (p. 417). This is

[8] I will speak of 'second representations' as opposed to 'second-order representations' because the second representations in cognitive decoupling are not representations of first-order representations, but are *copies* of representations. That a state is a copy of another state does not imply that it *represents* the state from which it was copied.

why a child wielding the whiffle-ball-bat sword can represent it as a sword, but simultaneously recognize that swinging it into his friend will not cut him. However, holding the secondary representation is cognitively costly since it involves suppression of the primary representation's influence on processes.[9] The child must suppress the primary representation of the bat, which includes that the bat is yellow, made of plastic, and lacks a guard that one would expect on a sword.

Nichols and Stich (2003) discuss cognitive decoupling in the context of pretense. Specifically, they argue that we need a new kind of mental state in addition to belief and desire to account for pretense: this they call 'supposing.' Supposing,' Nichols and Stich tell us, cannot be reduced to belief and desire. Thus, in addition to there being a 'belief box' and 'desire box' in their computational account of reasoning, there is a 'possible world' box. The possible world box only indirectly affects beliefs because both beliefs and suppositions both interact with inferences. One of the inference mechanisms is called 'updater,' and its function is to alter beliefs as circumstances change or novel beliefs are acquired. It searches for inconsistencies. It can also screen out beliefs that are inconsistent with those propositions currently in the possible world box. Crucially, though Stanovich make lite of this point, according to Nichols and Stich, the very same inference mechanism operates on supposing and beliefs (p. 29), supposing and beliefs do not differ in the type of content (the two are of the "*same logical form*" p. 32), and supposing and beliefs may combine in inferences (p. 32).

Since Stanovich has been the most explicit on how cognitive decoupling operates within a dual-process account, I will focus on Stanovich's accommodation of cognitive decoupling. Stanovich claims that the main mental ability measured by fluid intelligence is the mind's ability to sustain the ongoing simulations (Stanovich 2001, 2004), which is a function of the reflective

[9] Stanovich (2012) claims that "evolution has guaranteed the high cost of decoupling for a very good reason. As we were becoming the first creatures to rely strongly on cognitive simulation, it was especially important that we not become 'unhooked' from the world too much of the time. Thus, dealing with primary representations of the world always has a special salience" (50). But this seems odd. If it was important that our ancestors did not become 'unhooked' too much of the time, presumably the claim is that those who did become 'unhooked' too much of the time died off. The claim, it would seem, is that evolution found a way between two possibilities: one the one hand, creatures who did not think about concrete reality, and on the other hand creatures who only thought about concrete reality. Perhaps this is true, but further detail is needed as to why it is that having the *secondary* representation is so very useful, given that it is also cognitively costly. I am not convinced that an ability to sustain cognitive decoupling is evolutionarily advantageous. After all, the first philosopher was prone to falling into holes and being mocked by potential mates.

mind (Stanovich 2009, 2012). There are two kinds of cognitive decoupling on Stanovich's account. The first occurs when Type-1 processes are taken offline and stored away from Type-2 processing. Stanovich adopts the second kind of cognitive decoupling from developmental psychology. In this second kind of cognitive decoupling, a primary representation is copied and then used in abstract reasoning. It is this second kind of cognitive decoupling that concerns us most here. On Stanovich's account, the reflective mind issues a command to the algorithmic mind telling it to generate a copy of a representation. The algorithmic mind does so, and the reflective mind then uses that second representation in its abstract reasoning process (see chapter 1 and Figure 1.1 for more details on Stanovich's account). Language is one way to help with the second kind of cognitive decoupling because the second representation can be formed using natural language through inner speech (Carruthers 2006, Stanovich 2012). Stanovich then has an ontogenetic account of cognitive decoupling: humans engage in make-believe as children, which makes them able, as adults, to abstract away from the particulars of representation.

One might think that cognitive decoupling as the formation of a secondary representation only fits well with a multiple-system account of reasoning. This is not the case; Stanovich's account of cognitive decoupling is compatible with my one-system theory. The mere existence of a duality, such as primary and secondary representations, does not imply that there are two systems. A single system might represent both the world and the representation itself. However, there are independent reasons to be skeptical of Stanovich's account of cognitive decoupling.

Let me begin by pointing out a disanalogy between the decoupling in children's play and the decoupling in reasoning. Cognitive decoupling is supposed to be cognitively taxing, but children seem able to engage in make-believe play for extended periods of time, longer, in fact, than most of us can engage in abstract reasoning. The difference in ease of engagement between child make-believe play and adult abstract reasoning indicates that the decoupling involved in these two domains differs. Worse still, the child play (which is supposed to be training for adult abstract reasoning) is the easier of the two. If the ontogenetic story is right, then the reverse should be the case; children should at first find decoupling difficult, but it should become easier with time. Thus, although the ontogenetic story might be parsimonious (we engage in make-believe play as children, which enables us to do abstract reasoning as adults), it does not seem that the mechanism children use in pretend play is the same as the mechanism adults use in

abstract reasoning. If the two mechanisms are distinct, then we should not expect a strong correlation between the amount of time a subject engaged in pretend play and that subject's reasoning ability as an adult. Unfortunately, there have not been longitudinal studies looking for this link in particular. Perhaps this is due to the fact that such longitudinal studies would be difficult to run, since researchers would need to vary how much play time each child received, while controlling for all other variables.[10] However, there are some studies on children and play that suggested that play is important in childhood development in many domains. Bergen (2002) points to studies indicating that play is important to the development of theory of mind, mathematics readiness, literacy, linguistic ability, impulse control, representational competence, and problem-solving. The last is most relevant here. Smith and Dutton (1979) are responsible for demonstrating the correlation between free-play and problem-solving ability. However, Smith (1995) claims that he and Simon were unable to reproduce the results when they controlled for experimenter effects: in Smith and Dutton (1979) the same experimenter gave the play and training conditions and then tested the children on their problem-solving ability. Simon and Smith (1983, 1985) controlled for this by having two experiments: one for the play/training condition and another for the testing phase. Furthermore, I must point out that not all of the play was 'pretend play.' This casts more doubt as to whether these results should be taken to support the claim that the decoupling in pretend play and abstraction in reasoning are the same. Furthermore, it is unclear whether the problem-solving would have involved cognitive decoupling (children were required to join sticks together to retrieve a marble, which seems to be concrete, rather than abstract, reasoning). To conclude, there is no firm evidence that the mechanism responsible for cognitive decoupling in child play is the same in reasoning, since there is no firm evidence that increased play in childhood increases reasoning ability.

A second difficulty for these accounts of cognitive decoupling is that it is plausible that there are cases of cognitive decoupling that only involve primary representation. To see why, we must draw a distinction between *attending to specific details* of a primary representation and *forming a copy* (secondary representation) of a primary representation that includes all and only those specific details. First, notice that (in the case of abstract reasoning) the secondary

[10] Perhaps anecdotal evidence from the history of philosophy will help. John Stuart Mill was raised so as not to engage in any kind of fictitious play, and his powers of abstract reasoning as an adult were more than adequate.

representation needs to be devoid of the particulars of the primary representation. Thus, the second representation is not a perfect copy of the primary representation. Consider the cognitive decoupling involved in determining whether a valid argument with an unbelievable conclusion is valid. A good logician focuses solely on the *structure* of the argument, but there is no need for a secondary representation that is a copy of a representation of the argument. In fact, a secondary representation of the argument would only be useful over and above the first representation if the secondary representation was devoid of the content while preserving the argument's form. Indeed, we do not always teach our students to form secondary representations of arguments. We merely teach them to ignore the content while attending to the form. We then ask them if the *form* is valid or not. No secondary representation is *required*. Rather than needing a secondary representation which is stripped of some of the idiosyncrasies of the problem (e.g. being about a woman named Linda who majored in philosophy at Berkeley, for example), the reasoning system can operate on the primary representation alone by attending only to the relevant details. On this suggestion, cognitive decoupling does not necessarily involve the formation of a secondary representation; it might only involve operating on certain features (those deemed relevant) of a primary representations. The suggestion here is that, in solving (say) the Linda problem, there is no required intermediary stage in which one represents each of the putative answers as being of the forms 'A' and 'A and B.' That the two optional responses are of those forms is contained within the primary representation. The Linda case is paradigmatic of cognitive decoupling, but there is no need to introduce secondary representations.

In fact, there being a copy of a primary representation alone does very little to aid in the cognitive decoupling in cases like the Linda problem. As I pointed out above, the secondary representation is not identical to the first, since if the two were identical then any help having the secondary representation would be subsumed by having the first. The secondary representation aids in cognitive decoupling because it extracts relevant information and leaves out irrelevant information from the primary representation. Thus, it is the *selection* of whichever aspects of the primary representation will be kept in the secondary representation that is most important for cognitive decoupling, rather than the *formation* of the secondary representation. The key to

getting the Linda case right is to ignore the description when generating a response.[11] All that is relevant is the form of the two responses. Once this information is isolated, one only need recognize that a conjunction is never more likely than one of its conjuncts. Whether this is determined using a primary or secondary representation is irrelevant. Thus, there are cases of cognitive decoupling in which secondary representation is superfluous. If I am right in this, the startling consequence is that each token of cognitive decoupling does not necessarily require an instance of metacognition.

I am not claiming that secondary representations lack *any* utility, or that they are never utilized in cognitive decoupling; I am only claiming that there are cases where we need not introduce secondary representations to account for cognitive decoupling. However, there are cases in which a secondary representation would be helpful in cognitive decoupling. For example, when a subject needs to concentrate on certain properties of an argument or problem for a long time, it is useful to formulate a secondary representation that includes all and only those specific properties. Here the secondary representation is useful because the subject has less chance of allowing irrelevant variables to affect his or her response, since not only are they ignoring irrelevant details, those irrelevant details are not even contained in the secondary representation. Notice that the more spacious details are included in a primary representation, the more useful it will be to produce a secondary representation will be. That is, the more distracting the content and more difficult it is to extract the relevant details from an argument, the more useful it is to form a secondary representation that does away with those extraneous details. An example in which one consciously forms a secondary representation will help illuminate both cases. Suppose a subject is trying to determine whether the ontological argument for God's existence is valid. The subject might reasonably translate it into predicate logic to more accurately form his or her judgment. Because the subject needs to use the same propositions repeatedly, coming back to them after not thinking of them explicitly, and because all that is relevant to validity is the form of the argument, having a representation that captures only the structure of each proposition is useful. It is useful because the subject does not need to extract the structure of the propositions over and over again. This both saves time and aids in accuracy.

[11] I am assuming that 'picking out the salient details' is identical to 'ignoring the extraneous details.'

Cognitive decoupling might involve secondary representations, but secondary representations are not necessary for cognitive decoupling.

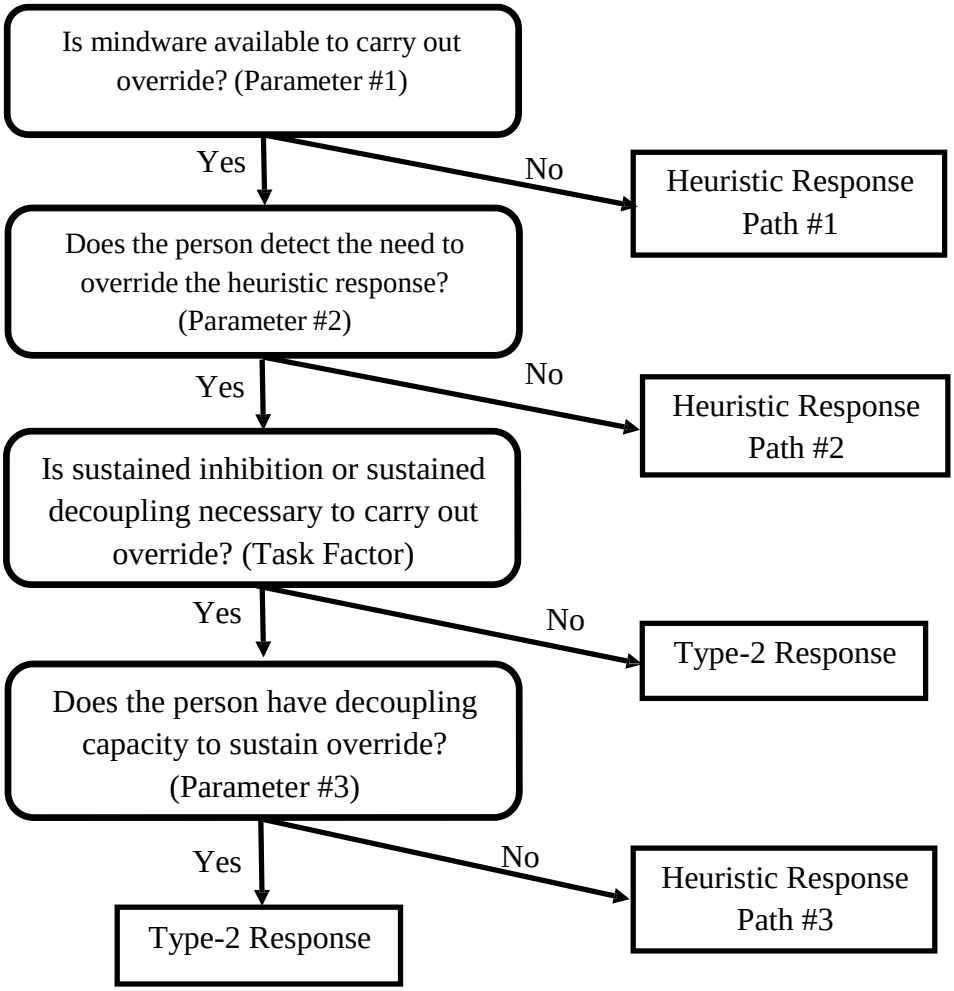
Suppose I am right in thinking that cognitive decoupling need not involve the copying of representations. Then cognitive decoupling does not 'fit well' with a certain conception of the relation between S1 and S2. Remember that cognitive decoupling is purported to be a function exclusively of S2 (or Stanovich's reflective mind). However, if cognitive decoupling can occur with only primary representations, then it is unclear why S1 would be unable, in principle, to do so. Furthermore, we now have a rough outline of how the one reasoning system moves from its concrete mode to its abstract mode through cognitive decoupling. Abstract reasoning is accomplished by attending to the relevant details of a primary representation, and may formulate a secondary representation devoid of irrelevant details in order to aid accuracy and speed of an abstract thinking process.

4. Malfunctions

Having examined how cognitive decoupling works on Stanovich's account and my account, we can now examine Stanovich's framework for conceptualizing individual reasoning errors, and I can offer my own rival taxonomy. One of Stanovich's most important contributions to the dual-process literature has been his emphasis on the importance of individual difference: although many subjects answer incorrectly in reasoning tasks, not all do. Furthermore, there are often tasks in which subjects get the wrong answer in different ways. A paradigmatic example of individual difference is the Wason Selection task where there are four main responses given by subjects, three of which are wrong. In most experiments involving the Wason selection task, no single response receives a majority. Stanovich rightly takes this to indicate that there is more than one malfunction at work in reasoning tasks. He (with Richard West[12]) improves upon Kahneman (2000) and Kahneman and Frederick's (2002) framework for when subjects deliver Type-1 or serial responses (the latter being generated by the algorithmic mind). According to Stanovich and West, the way that subjects fail to deliver the correct response depends upon whether there is mindware (i.e. the rules and procedures used to transform decoupled representations) available to override the Type-1 process. If not, then the subject will deliver a

[12] West co-authored a portions of Stanovich (2011).

Figure 5.1 Stanovich's Framework for Conceptualizing Individual Differences (Stanovich 2001, p. 143)



Type-1 response (Stanovich and West call this 'heuristic response path 1'). If the subject does have mindware available, then the subject might deliver the correct response. However, if the subject does not recognize that the Type-1 process or algorithmic response will not deliver the correct answer, the subject will not override the type-1 or algorithmic response, and thus will deliver a Type-1 or algorithmic response (Stanovich and West call this 'heuristic response path 2').[13] Supposing that the mindware for overriding the Type-1 or algorithmic response is available, and that the subject recognizes that an override is required, then (given that no

[13] Notice that, in order for the output of Heuristic Response Path 1 and Heuristic Response Path 2 to differ, it must be the case that Heuristic Response Path 2 uses the algorithmic mind, since if both used Type-1 processing, they would not differ in their responses.

decoupling is necessary), the subject will deliver a Type-2 response. If the process does involve sustained decoupling, then the subject will deliver a Type-2 response on the condition that they have the decoupling capacity. If they do not have the decoupling capacity (perhaps their working memory is too loaded, or they are under a time pressure), then they will deliver a Type-1 or algorithmic response (Stanovich and West call this 'heuristic response path 3'). The figure above (from Stanovich 2011) outlines the determination of whether a subject will deliver a Type-1 or Type-2 response.

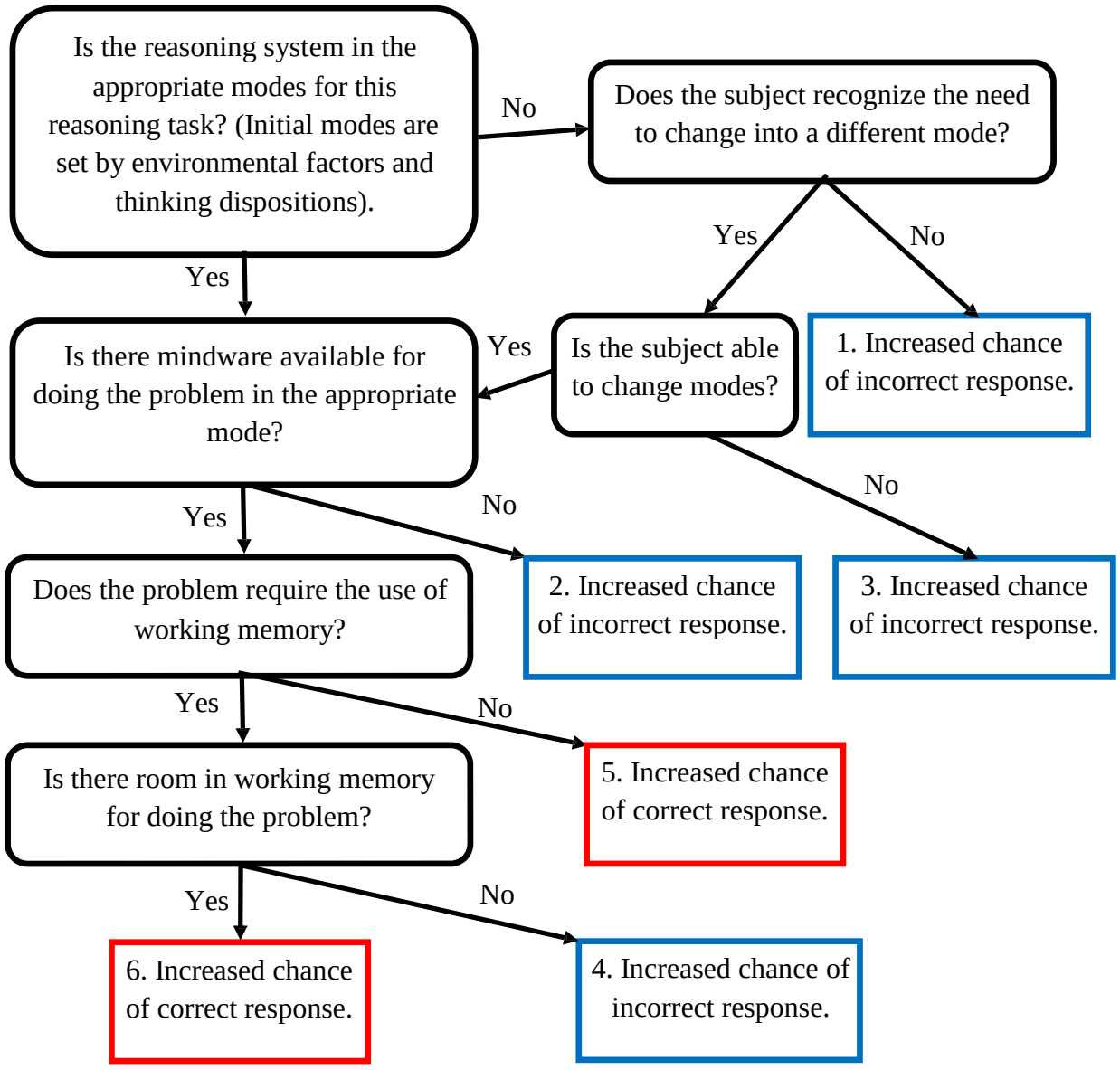
Notice that Stanovich's taxonomy is theory-laden with his own tripartite division of the mind. His theory's provision of this taxonomy is a virtue of his account of human reasoning. In order to compete with Stanovich's account, my one-system alternative should be able to provide an alternative to this dual-process taxonomy for thinking about individual difference. On the one-system theory, malfunctions are not merely due to a sub-optimal mode of reasoning being used. The reasoning system might not have the proper mindware to complete a certain problem. At the beginning of the semester, my logic students are not very good at determining whether an argument is valid, but by the end most are. This is because they need to acquire new mindware pertaining to formal logic. Consider a less controversial example: if you ask an eight-year-old, who has only just learned the concept of division, what 29837498/7 is, they will be unable to give you a correct answer. He or she needs to acquire the algorithm for doing long division. He or she needs new mindware.

Here follows my own framework for conceptualizing individual differences. On the one-system theory, responses vary because of differences in use of mode. So if a subject responds incorrectly, it is because they used a suboptimal mode (assuming the error is due to competence rather than performance). Thus, on my one-system account, a taxonomy for conceptualizing individual differences will depend on what determines the reasoning system's ability to operate in its optimal mode. I should point out that this framework, while intended to be compatible with my one-system theory, is not necessitated by it. The below framework can be altered while leaving my one-system theory intact. Furthermore, the framework is somewhat naïve because, as I have already said, the reasoning system does not necessarily remain in a definitive mode through the whole reasoning process. Rather, it can change modes as the reasoning process occurs. I have omitted these feedback arrows in my diagram below for the sake of simplicity.

First, it is important that the one reasoning system be in the appropriate modes. Again, which modes the reasoning system is in at the beginning of a reasoning problem depends upon environmental features and individual differences. If it is not in the optimal mode to begin with, then, does the subject recognize that the reasoning system should be in a different mode? If not (or if they are unable to operate in the most optimal mode), then the likelihood that the reasoning system will offer an incorrect response increases. If the reasoning system is operating in the optimal mode, does it possess the appropriate mindware within that mode for responding to the problem in question? If not, then the reasoning system may either try to complete the problem in the appropriate mode without the appropriate mindware, or it will operate in another mode. In either case, the likelihood of an incorrect response increases. Supposing that the correct mindware is available, does the problem require the use of working memory? If not, then the subject will likely give the correct response. If solving the problem does require the use of working memory, then is there enough space in working memory to carry out the reasoning process? If so, then the subject will likely offer the correct response. If not, then the subject will revert to some other mode to deliver a response.

Notice that, in my framework, I contrast 'increased chance of correct response' with 'increased chance of incorrect response.' Importantly, my claim is not that the likelihood of error is identical in paths 1-4 or 5-6. Indeed, these paths are misleadingly simple, as there are multiple suboptimal modes for many reasoning task. Furthermore, each of the paths on their own might differ from individual to individual, since, unlike Stanovich, I do not posit the existence of two kinds of processing. Thus, for example, path 1 for an individual problem might result in differences for different individuals because those individuals might have started in different modes due to differences in thinking dispositions. Also, whereas Stanovich and West contrast 'Type-2 response' with 'heuristic response path,' I contrast 'increased chance of correct response' with 'increased chance of incorrect response.' The reason I do not contrast 'correct' and 'incorrect' is twofold. First, just because the reasoning system operates in an optimal mode, with the correct mindware, and with enough time to complete the process, does not guarantee a correct response. Just as two-system theorists do not claim that S2 is infallible, I do not claim that the optimal mode with correct mindware is infallible. The subject might misunderstand the question, forget to carry a numeral, or misread his or her own handwriting when working out a

Figure 5.2 One-System Framework for Conceptualizing Individual Differences



logic or algebra problem. Perhaps these are not 'reasoning' flaws, since they are merely performance errors rather than competence errors. However, even if we do draw a firm distinction between performance and competence errors, both exist, and thus optimal reasoning does not guarantee a correct response to every reasoning problem. Second, I use 'increased chance of incorrect response' instead of 'incorrect response' because (like the two-system

theorist's thinking that S1 often delivers correct responses), a mode that is not optimal for delivering correct responses can still deliver a correct response.

5. Other one-system alternatives

My purpose in this section is to outline some alternatives to dual-process theory, which, for better or for worse, have not been very influential. I will outline two one-system theories and explain which aspects of these accounts are similar, and which aspects are different, from my own one-system theory.

Osman's (2004) one-system alternative is an extension of Cleeremans and Jimenez's (2002) dynamic graded continuum (DGC) theory of learning. On this connectionist account, quality of a representation lies along a continuum and depends on strength, stability, and distinctiveness, where "*strength* is defined as the amount and the level of activation of processing units,….*stability* is the length of time a representation remains active during processing,…[and] d*istinctiveness* refers to the discriminability of representations" (Osman 2004, p. 993). According to Osman, implicit, automatic, and explicit processing form a continuum. "Implicit reasoning involves making a set of abstractions or inferences without concomitant awareness of them" (p. 995). Subjects usually rely on implicit processing when they encounter novel reasoning problems. "Implicit reasoning is likely to result from situations where reasoners are unfamiliar with the task environment" (p. 996). In contrast to implicit (but not automatic) reasoning, subjects have awareness in *explicit reasoning*, and this awareness "can be expressed as declarative knowledge" (p. 995). Osman claims that explicit reasoning requires metacognition, since it involves thoughts about inferences. Finally, automatic reasoning is "deliberately acquired through frequent and consistent activation of relevant information that becomes highly familiarized" (p. 995). Importantly, subjects do not have control of the inferences, but do possess metaknowledge of them (p. 996). On her account, a procedure for solving a certain kind of reasoning problem may begin as explicit reasoning, but become automatic over time.

Osman (2004) explains that her "aim here is not to advance a new theory of reasoning but to provide a framework that can be used to assess dual-process theories and the evidence used to support their claims" (p. 993). Osman identifies four criteria for positing two systems: Criterion

S, individual differences, implicit versus explicit processing, and neuroanatomical difference. Interestingly, the DGC theory does not figure prominently in the section of her paper outlining alternative explanations of the data taken to support dual-process theory. Furthermore, reviewing her discussion of the data, it is unclear that individual differences and implicit versus explicit processing would count as evidence for the existence of two systems. She explains that although performance on reasoning tasks does correlate with cognitive ability, these results "can be interpreted as relating to differences in degree rather than in the kind of reasoning system used" (p. 1004). Also, positing implicit and explicit processes does not require the positing of two systems, since the data can "be interpreted as showing that participants vary as to the insight they have into their own reasoning" (p. 1004). On the other hand, she seems to think Criterion S and neuroanatomical differences would provide good evidence for two systems, but are currently unfulfilled.

I am sympathetic to DCG, and my account of human reasoning is similar to it in some ways. Namely, like DCG, my account allows for a spectrum between certain opposing properties, whereas dual-process theories posits a pair of exhaustive alternative. On my account, the automatic/controlled distinction is a matter of degree, and, like Osman, I suggest (chapter 4) that processes that begin as controlled might become automatic over time. Osman posits 'implicit' processing and claims that it is on a continuum with the automatic and explicit processing. I agree with Osman in thinking that implicit and automatic might come apart, contra dual-process theory. However, since, according to Osman, automatic (in her sense) processes do not become implicit over time, it is odd to put these three on a continuum. Thus, it is better to oppose implicit and explicit (i.e. unconscious and conscious), on the one hand, and automatic and controlled on the other. The automatic/controlled distinction and the implicit/explicit distinction are both continuums, and, though related, are distinct. My account makes this clear. To return to my soundboard analogy, whereas my account posits one slide for automatic/controlled and another for implicit/automatic, Osman posits only one slide.

Kruglanski and Gigerenzer's (2011) sketch of a unified theory of reasoning consists of six claims:

1. Intuitive and deliberative judgments are both based on rules, and the very same rules can underlie both.

2. Kruglanski and Gigerenzer claim that there are at least four factors that determine which rule a subject selects for solving a problem. The task itself and individual memory both "constrain the set of applicable rules," and then "individual processing potential and (perceived) ecological rationality of the rule…guide the final selection" of the rule (p. 98, see p. 102-103). The rule selection process is not "deliberative" or conscious (p. 100).
3. Rule conflict occurs when subjects perceive two or more rules having equal ecological rationality. "In such cases, proper application of a given rule may suffer interference from other competing rules" (p. 98, see p. 104).
4. Individual differences in cognitive ability "influence speed and accuracy with which a rule is executed" (p. 98). However, a rule's being difficult or easy to apply does not correlate with a rule's being intuitive or deliberative.
5. The more difficult a rule is to apply, the more processing power is required for applying it. Thus, when cognitive energy is limited, subjects will only apply easy rules. However, when subjects have lots of cognitive energy, ecological rationality (rather than ease) determines rule use.
6. "The accuracy of both deliberate and intuitive judgments depends on the ecological rationality of the rule for the given class of problems" (p. 98).

Kruglanski and Gigerenzer outline ten heuristics that are "likely in the adaptive toolbox of humans" (p. 101). They go on to offer alternative explanations of the data from the heuristics and biases literature using these heuristics and identifying how subjects decide which rule to use in given tasks.

I agree with Gigerenzer in thinking that heuristics should be construed as rule-based. However, I think Gigerenzer goes wrong in doing away with associations altogether. I am skeptical that Gigerenzer can explain all the data from the heuristics and biases literature using only the somewhat simple heuristics he lists in his article. To be sure, Gigerenzer can always add more rules, and, to some extent, does so in explaining how subjects determine which rule they will use. Whereas it is unclear how to test dual-process theories, Gigerenzer rightly points out that his account is testable. Note that what is testable is which specific heuristics subjects use. However, because Gigerenzer denies the existence of associations, he must posit complex rules, perhaps unnecessarily complex. One might also question whether subjects are following the rules Gigerenzer outlines, rather than merely conforming to them. In reply, Kruglanski and Gigerenzer claim that rule conforming is rigid in nature, while subjects' use of rules is flexible (i.e. they can be learned, unlearned, forgotten, and retrieved) (p. 100). However, it is unclear why this is a good criterion for following rather than conforming to a rule. It may, as a result, be simpler to posit the existence of associations in addition to rule-based processes.

6. Empirically distinguishing one-system and default-interventionist dual-process theories

In chapter 3, I argued that any one-system theory will be incompatible with the existence of SCBs arising from reasoning. I argued that this gives us a clear way to empirically distinguish parallel two-system theories from one-system alternatives. It is unclear whether default-interventionists would allow for the existence of SCB. If they do not, then one might worry that one-system alternatives cannot be empirically distinguished from default-interventionist dual-process theories. Given Evans and Stanovich's endorsement of this position, the worry cannot be taken lightly. In reply to this challenge, I will suggest a way to empirically distinguish *my* one-system alternative from default-interventionist dual-process theories. While I focus on Evans and Stanovich's (2013a) most recent accounts, these points apply equally to any account that takes automaticity and working-memory as central to the Type-1/Type-2 or S1/S2 distinctions. After outlining some differences in prediction, I argue that experiments from De Neys (2006) confirm predictions generated by my theory over default-interventionism. Next, I suggest how Stanovich and Evans might amend their account to accommodate these findings. However, I argue that these amendments strip dual-process theory of any substantive claim.

How much working memory do we use for reasoning tasks? According to Stanovich and Evans (2013a), there are two kinds of reasoning processes: those that are mandatory (or 'autonomous' as they call them)[14] and those that involve working memory. Thus, they posit two modes of control. So for any given reasoning task, we should expect subjects to use one of the two levels of control. However, on my account, I allow for a gradation. There are not two modes of use of working memory, but a smooth gradation. Subjects can use more of less of their working memory in various tasks. To use my sound mixer analogy again, automatic/controlled is a slide, not a switch. Analogously, Pim Haselager (2014) suggests that according to dual-process theory, reasoning is like a car that has two gears. Haselager suggests that we need an account of reasoning according to which control is like a car with many gears (though he is unsure whether we need 5 or 500). My account provides just such an account. (Exactly how many degrees of control there are, and whether they are analogue or digital is an empirical question, and one I will not take up here).

[14] One might object to Evans and Stanovich's contrasting mandatory with working-memory involving on grounds that they are not mutually exclusive categories. In this section, I will assume that autonomous processes do not involve working-memory.

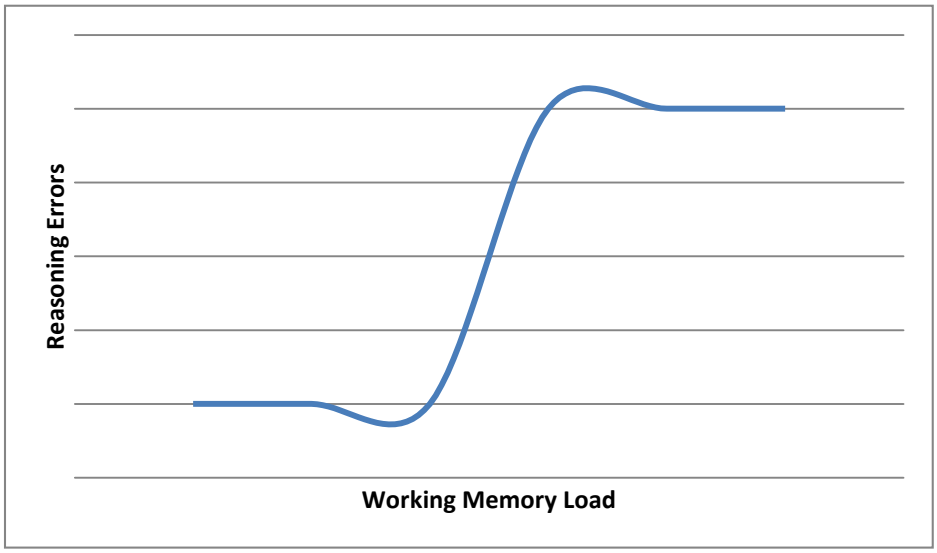
These theoretical differences lead to two different predictions concerning interference under cognitive load. We know that putting a load on working memory increases many biases (especially belief bias) (Evans and Curtis-Holmes 2005, De Neys 2006). Of course, the amount of load put on working memory is not all or nothing. Researchers can put greater and lesser cognitive load on subjects. Dual-process theory, it would seem, should predict a threshold where the working memory load becomes too great for subjects to engage in Type-2 reasoning and are therefore forced to use only their Type-1 reasoning. To use Stanovich's taxonomy for individual differences in reasoning errors, even subjects who avoid the first two heuristic paths will end up taking the third heuristic path. Thus, as researchers increase load on working-memory, there should be a jump at some point in reasoning errors. The jump will occur at the point where working memory load is high enough to prevent subjects from using their Type-2 reasoning for the purpose of solving the reasoning tasks at hand. That is, at the point where subjects can, at best, use heuristic path 3. Furthermore, when subjects are forced not to reason using their Type-2 processing, they must use their Type-1 processing, and using Type-1 processing, since it does not use working memory, will not be affected by increased cognitive load. Thus, once subjects' working memory is taxed to the point that they are forced to use their Type-1 processing, adding additional load on working memory should not increase reasoning errors. Thus, Stanovich and Evans's accounts predict that the relation between the amount of reasoning errors and the load on working memory will resemble a sigmoid function (see Figure 5.2).

However, my account predicts no such threshold. If there is only one reasoning system that can operate more or less automatically or in a controlled manner, then there should be a smooth relation between how much load is put on working memory and how many errors are made in reasoning tasks. I am agnostic as to whether the relation between cognitive load and reasoning errors is exponential or linear. What I am committed to is that, if my account is right, then the relation between cognitive load and reasoning errors is not expressed by a sigmoid function.

One might object to these predictions on grounds that greater control does not necessarily imply better reasoning (Carruthers 2013a, Evans and Stanovich 2013a). Indeed, that greater control does not imply greater accuracy helps my case that the properties on the Standard Menu cross-cut one another. However, there are some reasoning problems for which greater control is

closely correlated with better reasoning. Specifically, correctly responding to syllogistic reasoning tasks requires working memory. We know this because we know that increased cognitive load increases belief bias (De Neys 2006).

Figure 5.3: Prediction on Default-Interventionism



There is another objection. Stanovich (2009, 2011) posits the existence of three minds, two of which engage in Type-2 reasoning. The algorithmic mind and reflective mind both engage in Type-2 reasoning, but the Algorithmic mind uses (generally) far less working memory than the reflective mind. On Stanovich's account, we might expect two steps in the graph instead of one—Type-1 reasoning involves no working memory, Type-2 reasoning as carried out by the algorithmic mind involves a certain degree of working memory, and Type-2 reasoning as carried out by the reflective mind involves a yet greater degree of working memory.

Some of Evans's work suggests that the amount of working memory used in a process will vary in degree. Evans posits a process wherein subjects decide how much cognitive energy they will use in a given reasoning problem (in 2009 he calls this 'Type- 3' processing, but he later drops this terminology). Before engaging in a reasoning task, subjects must determine how much cognitive energy they will devote to a given reasoning problem, and Evans suggests that this determination itself is made by reasoning. However, the determination of how much working memory to use on a reasoning problem is itself a reasoning process. Indeed, it is a

reasoning process by hypothesis on Evans's account, since the deliberation as to how much cognitive energy one should use is determined by Type-2 processing, which is (by hypothesis) reasoning. Since the determination of how much cognitive energy one will expend is itself a reasoning process, there will have to be a further process to determine how much cognitive energy the subject will devote to it. Thus, we have a regress. This is a deep flaw in Evans's attempt to allow for difference in degree, as well as kind, of working-memory involvement.

Figure 5.4: The Relation between Cognitive Load and Reasoning Errors (from De Neys 2006)

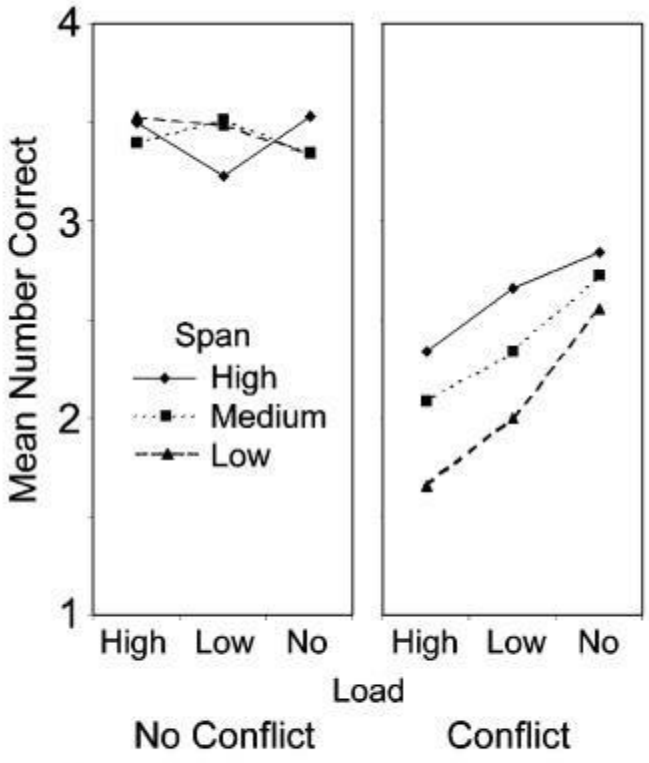


Fig. 2. Reasoning performance of the high-, medium-, and low-span groups as a function of executive load. Results are shown separately for conflict problems, in which the logical validity of the conclusion conflicted with its believability, and no-conflict problems, in which the logical validity and believability of the conclusion were consistent.

There is some clear evidence in favor of my prediction. De Neys (2006) tested individuals under three conditions: no load on working memory, medium load on working memory, and high load on working memory. He used a dot memory task for the medium and high loads put on working memory (see chapter 2 of this dissertation for details). As predicted by

the one-system theory, "performance decreased linearly with increasing secondary-task load" (p. 431). De Neys only takes this to indicate that "executive working memory resources" are required for proper reasoning in conflict cases (p. 431). (While De Neys takes his results to support the existence of two reasoning systems, see chapter 2 of this dissertation for an argument that his results are compatible with one-system alternatives). Admittedly, De Neys (2006) only tested subjects under three levels of cognitive load. My case would be stronger if we saw this linear correlation between reasoning errors and many degrees of load on working memory. Although I do not know of any further studies on syllogistic reasoning and degree of load put on working memory, experimenters could test subjects under more degrees of cognitive load.

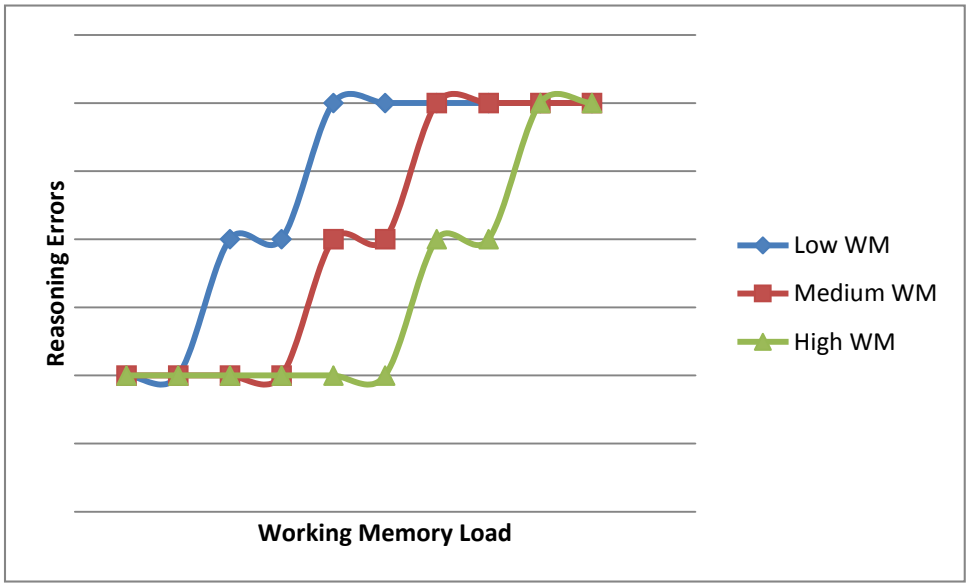
Stanovich might reply that subjects in the medium cognitive load condition used their algorithmic mind while those under no cognitive load used their reflective mind. There are some problems with this proposal. First, the algorithmic mind is of no help whatsoever if the appropriate mindware is not available. The subjects De Neys used were supposed to be syllogistically naïve; they did not have the relevant mindware. Thus, they would either be forced into reflective thinking or Type-1 thinking.[15] Therefore, Stanovich's theory should predict that there would be a sharp turning point. There is a second problem for Stanovich even if subjects were able to use their algorithmic minds. De Neys divided subjects into three groups according to the amount of working memory capacity. We should expect that the threshold for switching to Type-1 processing varies with working memory capacity. As such, the low capacity subjects should have a lower threshold than the high capacity subjects, and the medium capacity subjects should have something in between (see Figure 5.5).

This alternative prediction is also not consistent with the results in De Neys (2006). We do not find each group (high, medium, and low working-memory capacity) differing in their 'switching point.' Each group's reasoning errors increased under similar load conditions. These findings strongly suggest that there are not three levels of working-memory involvement, the switching points of which vary by individual. Instead, Stanovich would have to say that there are three levels of working-memory involvement, the switching point of which varies by individual

[15] Which they use would depend on how much working memory they had available, combined with their thinking dispositions such as how willing they were to engage in difficult reasoning tasks. However, *why* they are forced into one type of thinking over another is not relevant here. All that is relevant is *that* they are forced into one position or another.

and the extent to which the kind of processing at each level outputs the correct response also varies individually.

Figure 5.5: Amended Prediction on Default-Interventionism:



Both Stanovich and Evans suggest that although Type-2 processing uses working memory, this does not imply that Type-2 processing is 'all or nothing.' In the medium load on working memory conditions, subjects might be able to use their Type-2, but not as fully as those who have no load on working memory. Furthermore, Type-2 reasoning might have many degrees of control. Thus, dual-process theory is compatible with a linear relation between load on working memory and reasoning errors. This position is consistent, but threatens to strip dual-process theory of its substantial claims. Dual-process theory, being an empirical theory, requires that the two kinds of processes be *empirically*. Stanovich and Evans (2013a) tell us that these kinds are distinguished in that Type-2 processes use working memory, whereas Type-1 do not. However, if dual-process theorists adopt the suggestion outlined above, then Type-2 processing can use more or less working memory. Now, if reasoning processes are to be distinguished into kinds by the amount of working memory they involve, then it would seem that Type-2 reasoning will likely be split into more types. We will need a type of reasoning for medium use of working memory, another for high use of working memory, and perhaps many more for between. The

worry here is that the dual-process theorist now is forced into adopting an implausible account of reasoning processes, one which is cluttered. The dual-process theorist might reply that there is a qualitative distinction between involving *no* working memory and involving *some* working memory (even if it is very little), but there is only a quantitative difference between using *very little* working memory and using *a lot* of working memory. The problem with this suggestion is that it trivializes dual-process theory. Any view of the mind that posits working memory (but which does not claim that working memory is required for every reasoning task) would then be committed to dual-process theory.

7. Conclusion

I have argued for a conception of reasoning as engaging in inferences, where inference drawing is understood as a relation between two or more assertions such that the subject judges (perhaps implicitly) that there is a justifying relation obtaining between the two. I have suggested that the human reasoning system can operate unconsciously or consciously, concretely or abstractly, mandatorily (in the sense that once it is started it cannot be stopped) or in a controlled manner, and inductively or deductively. I have outlined how the one reasoning system can move between concrete and abstract reasoning—which is perhaps the most important property pair to consider in some of the paradigmatic reasoning errors. I have argued that abstract reasoning need not require second-order representation and have outlined when the reasoning system might use second-order representations. I have offered a rival to Stanovich's taxonomy of individual difference. Finally, I have offered empirical evidence for my one-system theory over default-interventionist dual-process theory.

Bibliography

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120,* 3-19.

Apperly, I. (2010). *Mindreaders: the cognitive basis of 'theory of mind.'* Psychology Press.

Apperly, I. A., and Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states?. *Psychological Review*, 116, 953.

Aizawa, K., and Gillett, C. (2009). The (multiple) realization of psychological and other properties in the sciences. *Mind & Language*, 24, 181-208.

Bergen, D. (2002). The role of pretend play in children's cognitive development. *Early Childhood Research & Practice*, *4*, 2-13.

Bigelow, J., and Pargetter, R. (1987). Functions. *The Journal of Philosophy*, 181-196.

Bourgeois-Gironde, S., and Van Der Henst, J. B. (2009). How to open the door to System 2: Debiasing the Bat-and-Ball problem. *Rational Animals, Irrational Humans*, 235-252.

Bratman, M. E. (1987). *Intentions, plans, and practical reason.* Cambridge, MA: Harvard University Press.

Bratman, M. E. (1999). *Faces of intention.* Cambridge: Cambridge University Press.

Brooks, R. (1991). Intelligence without reason. *Artificial Intelligence,* 47, 139-159.

Brooks, R. (1999). *Cambrian intelligence*. Cambridge: MIT Press.

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

Boyd, R. (1999). Homeostasis, species, and higher taxa. In Wilson, R. (Ed.), *Species: new interdisciplinary essays* (pp. 141–185). Cambridge, MA: MIT Press.

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

Carruthers, P. (2009) An architecture for dual reasoning. In Evans, J. St. B. T. and Frankish, K. (Eds.), *In two minds: Dual processes and beyond* (pp. 109–27). Oxford University Press.

Carruthers, P. (2011). *The opacity of mind: an integrative theory of self-knowledge*. Oxford University Press.

Carruthers, P. (2013a) The fragmentation of reasoning. In P. Quintanilla (Ed.), *La coevolución de mente y lenguaje: Ontogénesis y filogénesis*. Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú.

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

Carruthers, P. (2013c). The distinctively-human mind: the many pillars of cumulative culture. In Hatfield, G. and Pittman, H. (Eds.), *The evolution of mind, brain, and culture*. Penn Museum Press.

Chemero, A. (2000). Anti-representationalism and the dynamical stance. *Philosophy of Science*, 625-647.

Cheney, D. and Seyfarth, R. (1985). Social and non-social knowledge in vervet monkeys. *Philosophical Transactions of the Royal Society B vol. 308 no. 1135 187-201.*

Cohen, J. (1992).An essay on belief and acceptance. Oxford: Clarendon Press.

Cowey, A. and Stoerig, P. (1997) Visual detection in monkeys with blindsight. *Neuropsychologia* 35, 929–939

Craigie, Jillian. (2011). Thinking and feeling: Moral deliberation in a dual-process Framework. *Philosophical Psychology* 24:1, 53-71.

Cummins, R. (1975). Functional explanation. *Journal of Philosophy*, 72, 741-764.

Cummins, R. C. (2000). "How does it work" versus "what are the laws?": Two conceptions of psychological explanation. In Keil, F. and Wilson, R. A. (Eds.), *Explanation and cognition* (pp. 117-145). MIT Press.

Darlow, A. L., and Sloman, S. A. (2010). Two systems of reasoning: Architecture and relation to emotion. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 382-392.

Dawkins, R. (1982). *The extended phenotype.* New York: Oxford University Press.

De Houwer, J. (2014). A Propositional of Implicit Evaluation. *Social and Personality Psychology Compass* 8(7): 342-353.

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17,* 428-433.

De Neys, W., & Goel, V. (2011). Heuristics and biases in the brain: Dual neural pathways for decision making. In O. Vartanian & D. R. Mandel (Eds.), *Neuroscience of Decision Making* (pp.125-141). Hove, UK: Psychology Press.

De Neys, W., Rossi, S., and Houdé, O. (2013). Bats, balls, and substitution sensitivity: cognitive misers are no happy fools. *Psychon Bull Rev*iew 20(2):269-73.

Dennett. (1978). *Brainstorms: Philosophical essays on mind and psychology*. Montgomery, Vt.: Bradford Books.

Dennett, D. (1987). *The intentional stance.* Cambridge, MA: MIT Press.

Dennett, D. (1988). Précis of *The intentional stance. Behavioral and Brain Sciences,* 11, 493-544.

Dennett, D. (1995) The path not taken. *Behavioral and Brain Sciences,* 18, 252–253.

Devine, P. G. (1989). Stereotypes and prejudice: their automatic and controlled components. *Journal of personality and social psychology*, *56*(1), 5.

Dienes, Z. and Perner, J. (1999) A theory of implicit and explicit knowledge. *Behavioral and Brain Science*, 22, 735–808.

Duhem, P. (1906). *The aim and structure of physical theory.* Translated by Philip Wiener. Princeton: Princeton University Press.

Evans, G. (1982). *The Varieties of Reference*. New York: Oxford University Press.

Evans, J. S. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, 75, 451–468.

Evans, J. St. B. T. (1996). Deciding before you think: relevance and reasoning in the selection task. *British Journal of Psychology*, 87, 223–40.

Evans, Jonathan St. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Science* 7, 454-459.

Evans, Jonathan St. B. T. (2008). Dual-Processing Account of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, 59, 255–278.

Evans, J.St.B.T. (2009b) Introspection, confabulation and dual-process theory. Behavioral and Brain Sciences, 2, 142-143. (Commentary on Carruthers).

Evans, Jonathan St. B. T. (2010a). *Thinking twice: two minds in one brain*. Oxford and New York: Oxford University Press.

Evans, J. St. B. T. (2010b). Intuition and reasoning: A dual-process perspective. *Psychological Inquiry*, 21, 313–326.

Evans, J. St. B. T. (2011b). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review*, 31, 86–102.

Evans, J. St. B. T. (2012). Dual-process theories of reasoning: Facts and fallacies. In K. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 115–133). New York, NY: Oxford University Press.

Evans, J. S. B. T. (2013a). Selective processes in reasoning. In Evans, J. S. B. (Ed.) *Thinking and reasoning (psychology revivals): Psychological approaches* (pp. 135-163). Psychology Press.

Evans, J. S. B. (Ed.). (2013b). *Thinking and reasoning (psychology revivals): Psychological approaches*. Psychology Press.

Evans, J. S. B. T, Barston, J., and Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition* 11, 295-306.

Evans, J. S. B. T., and Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, 11, 382-389.

Evans, J. St. B. T., and Frankish, K. (Eds) (2009). *In two minds: Dual processes and beyond*. New York, NY, US: Oxford University Press.

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

Evans, J. St. B. T., and Stanovich, K. E. (2013a). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science* 8.3: 223-241.

Fiala, B., Arico, A., and Nichols, S. (2011). On the psychological origins of dualism: Dual-process cognition and the explanatory gap. In Slingerland, E., and Collard, M. (Eds), *Creating consilience: Issues and case studies in the integration of the sciences and humanities*. New York: Oxford University Press.

Fodor, J. (1968). *Psychological explanation*. New York: Random House.

Fodor, J. (1975), *The language of thought*. New York: Cromwell.

Fodor, J.A. (1983), *The modularity of mind*, MIT Press.

Fodor, J.A. (1987). Modules, frames, fridgeons, sleeping dogs, and the music of the spheres. In Pylyshyn, Z. (Eds.) *The robot's dilemma: The frame problem in artificial intelligence*. Norwood, NJ: Ablex.

Fodor, J. A., and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.

Frankish, K. (2004). *Mind and supermind*. Cambridge University Press.

Frankish, K. (2010). Dual-process and dual-system theories of reasoning. *Philosophy Compass*, 5, 914-926.

Frankish, K. (2012). Dual systems and dual attitudes. *Mind and Society* 11, 41-51.

Frankish, K., and Evans, J. St. B. T. (2009). The duality of mind: An historical perspective. In Evans, J. St. B. T., and Frankish, Keith (Eds), *In two minds: Dual processes and beyond*, (pp. 1-29). New York, NY, US: Oxford University Press.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of economic perspectives*, 25-42.

Gallistel, C. R., and Gibbon, J. (2001). Computational versus associative models of simple conditioning. *Current Directions in Psychological Science, 10,* 146-150.

Gallistel, C. R., and King, A. P. (2009). *Memory and the computational brain: Why cognitive science will transform neuroscience.* Singapore: Wiley-Blackwell.

Gigerenzer, G. (2000). *Adaptive thinking*. Oxford University Press.

Gigerenzer, G. (2010). Personal reflections on theory and psychology. *Theory & Psychology,* 20, 733–743.

Gigerenzer, G., and Regier, T. (1996). How do we tell an association from a rule? Comment on Sloman (1996). *Psychological Bulletin*, 119, 23–26.

Gould, S. J. (1991). *Bully for brontosaurus: Reflections in natural history*. New York: Norton & Company.

Goel, V. (2005). Cognitive Neuroscience of Deductive Reasoning. In Holyoak, K. J. and Morrison, R. G. (Eds.), *Cambridge handbook of thinking & reasoning*. Cambridge, MA: Cambridge University Press.

Goel, V., Buchel, C., Rith, C., and Olan J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *NeuroImage* 12:504–14

Goel, V., Dolan R. J. 2003. Explaining modulation of reasoning by belief. *Cognition*, 87, 11–22.

Greene, J. D., Nystrom, L. E., Engel, C. L., Darley, J. M., and Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44, 389–400.

Greene, J. D. (2008). The secret joke of Kant's soul. In Sinnott-Armstrong, W. (Ed.), *Moral psychology: Vol. 3. The neuroscience of morality: Emotion, brain disorders, and development* (pp. 35–79). Cambridge, MA: MIT Press.

Greene, J. D. (2009). Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology*, 45, 58–84.

Griffen, Z., and Ferreira, V. (2006). Properties of Spoken Language production. In *Handbook of psycholinguistics 2nd Edition* ed. Trazler and Gernsbacher. Amsterdam: Elsevier.

Griggs, R. A., and Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, 73, 407-420.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814–834.

Haselager, P. (2014). Cognitive science and common sense. ESPP 2014 Invited Symposium. Organized by Pietro Pertoni.

Hawthorne, J. (2005). The case for closure. In Steup, M. and Sosa, E. (Eds.), *Contemporary debates in epistemology* (pp. 26-42). Blackwell.

Heil, J. (2003). *From an ontological point of view*. Oxford University Press.

Hollander, M. A., Gelman, S. A., and Star, J. (2002). Children's interpretation of generic noun phrases. *Developmental Psychology*, 38, 883.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 90-98.

Kahneman, D., and Tversky, A. (1996). On the reality of cognitive illusions. A reply to Gigerenzer's critique. *Psychological Review*, 103, 582-591.

Kahneman, D. (2002). Maps of bounded rationality: A perspective on intuitive judgment and choice. In Frangsmyr, T. (Ed.), *Nobel Prizes 2002: Nobel Prizes, presentations, biographies, & lectures* (pp. 416–499). Stockholm, Sweden: Almqvist & Wiksell.

Kahneman, D., and Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In Gilovich, T., Griffin, D., and Kahneman, D. (Eds.), *Heuristics and biases*, (pp. 49–81). New York: Cambridge University Press.

Kahneman, D. (1994). New Challenges to the rationality assumption. *Journal of Institutional and Theoretical Economics, 150/1*, 18-36.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

Kennett, J., and Fine, C. (2008). Internalism and the evidence from psychopaths and 'acquired sociopaths.' In W. Sinnott-Armstrong (Ed.), *Moral psychology: Vol. 3. The neuroscience of morality: Emotion, brain disorders, and development* (pp. 173–190). Cambridge, MA: MIT Press.

Keren, G. (2013). A tale of two systems a scientific advance or a theoretical stone soup? Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science* 8.3: 257-262.

Keren, G., and Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science,* 4, 533–550.

Köhler, W. (1927). *The mentality of apes.* New York: Harcourt Brace.

Krachun, C. Call, J., and Tomasello, M. (2009). Can chimpanzees discriminate appearances from reality? *Cognition,* 112, 435-450.

Kriegel, U. (2012). Moral motivation, moral phenomenology, and the alief/belief distinction. *Australasian Journal of Philosophy*, 90:3, 469-486

Kruglanski, Arie W. (2013). Only one? The default interventionist perspective as a unimodel—Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science* 8.3: 242-247.

Kunda, Z. (1999). *Social cognition: Making sense of people.* MIT Press.

Leslie, A. M. (1987). Pretense and representation: The origins of theory of mind. *Psychological review*, *94*(4), 412.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind.' *Trends in cognitive sciences*, *8*(12), 528-533.

Leslie, S. J. (2007). Generics and the structure of the mind." *Philosophical Perspectives*, 21, 375-403.

Leslie, S. J. (2008). Generics: Cognition and acquisition. *Philosophical Review*, 117, 1-47.

Leslie, S. J. (2012). Generics Articulate Default Generalizations. *Recherches Linguistiques de Vincennes: New Perspectives on Genericity at the Interfaces*, 41, 25-45.

Levinson, S. C. (1995). Interactional biases in human thinking. In E. Goody (Ed.), *Social intelligence and interaction* (pp. 221-260). Cambridge: Cambridge University Press.

Lewis, D. (1973). *Counterfactuals.* Oxford: Basil Blackwell.

Lewis, David. (1986). Causal Explanation. In *Philosophical Papers: Volume II*. Oxford: Oxford University Press.

Lieberman, M. D. (2007). The X- and C-systems: The neural basis of automatic and controlled social cognition. In E. Harmon-Jones & P. Winkelman (Eds.), *Fundamentals of social neuroscience* (pp. 290–315). New York: Guilford Press.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology,* 37, 435-442.

MacLeod, C.M. (1991). Half a century of research on the Stoop effect: An integrative review. *Psychological Bulletin, 109* 163-203.

Mallon, Ron & Nichols, Shaun (2011). Dual processes and moral rules. *Emotion Review,* 3, 284-285.

Mandelbaum, Eric (2013). Thinking is Believing. Inquiry 57 (1):55-96.

Marslen-Wilson, W. (1973). *Speech shadowing and speech perception.* Ph.D. Thesis, MIT.

Marslen-Wilson, W. and Tyler, L. (1981). Central processes in speech understanding. *Philosophical Transactions of the Royal Society, B* 295: 317-322.

Martin, C. B. (2008). *The mind in nature*. Oxford University Press.

Masicampo, E. J., & Baumeister, R. F. (2008). Toward a physiology of dual-process reasoning and judgment: Lemonade, willpower, and expensive rule-based analysis. *Psychological Science, 19*(3), 255-260.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. New York: Psychology Press.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *The Journal of Neuroscience, 16*(16), 5154-5167.

Millikan, R. (1984). *Language thought and other biological categories: New foundations for realism.* Cambridge, MA: MIT Press.

Mugg, J. (2013). Why dispositions are not higher-order properties. *Proceeding of the Gesellschaft für Analytische Philosophie e.V*: 104-110.

Mumford, S. (1998). *Dispositions.* Oxford University Press.

Nagel, Jennifer. (2011). The psychological basis of the Harman-Vogel Paradox.' *Philosophers' Imprint* 11, 1-28.

Neander, K. (1991). The teleological notion of 'function.' *Australasian Journal of Philosophy*, 69:4, 454-468.

Nisbett, R. E., and T. D. Wilson. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84:231–59.

Osherson, D., Smith, E.E., Wilkie, O., Lopez, A., and Shafir, E. (1990). Category-based induction. *Psychological Review,* 97: 185-200.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bullentin & Review 11,* 988-1010.

Pollock, J. L. (1991). OSCAR: A general theory of rationality. In J. Cummins & J. L. Pollock (Eds.), *Philosophy and AI: Essays at the interface* (pp. 189-213). Cambridge, MA: MIT Press.

Prasada, S., and Pinker, S. (1993). Generalization of regular and irregular morphological patterns. *Lang. Cogn. Proc*. 8, 1–56

Prior, E. W., Pargetter, R., and Jackson, F. (1982). Three theses about dispositions. *American Philosophical Quarterly*, 251-257.

Polger, T. W. and Shapiro, L. (2008). Understand the dimensions of realization. *Journal of Philosophy*, 105, 213-222.

Putnam, H. (1967). Psychological predicates. In W.H. Capitan and D.D. Merrill (Eds.) *Art, mind, and religion* (pp.37-48). Pittsburgh: University of Pittsburgh Press.

Pylyshyn, Z. W. (1984). *Computation and cognition*. Cambridge, MA: MIT press.

Pylyshyn, Z. W. (Ed.). (1987). *The robot's dilemma: The frame problem in artificial intelligence* (Vol. 4). Norwood, NJ: Ablex.

Reber, A. S. (1993). *Implicit learning and tacit knowledge*. New York: Oxford University Press.

Revlin, R., Leirer, V., Yopp, H., & Yopp, R. (1980). The belief-bias effect in formal reasoning: The influence of knowledge on logic. *Memory & Cognition*, 8, 584-592.

Richeson, J. A., and Shelton, J. N. (2007). Negotiating interracial interactions: Costs, consequences, and possibilities. *Current Directions in Psychological Science*, 16(6), 316–320.

Samuels, R. (2009). The magical number two, plus or minus: Dual-process theory as a theory of cognitive kinds. In J. St. B. T. Evans and K. Frankish (Eds.). *In two minds: Dual processes and beyond* (pp. 129-146). New York: Oxford University Press.

Shafir, E., Smith, E. E., and Osherson, D. (1990). Typicality and reasoning fallacies. *Memory & Cognition*, 18, 229-239.

Shapiro, L. A. (2000). Multiple realizations. *Journal of Philosophy* 97, 635-654.

Shapiro, L. A. (2004). *The mind incarnate*. MIT Press: Cambridge.

Shoemaker, S. (1980). Causality and Properties. *Time and Cause: Philosophical Studies Series in Philosophy*, 19, 109-135.

Simon, T. and Smith, P. K. (1983). The study of play and problem solving in preschool children: Have experimenter effects been responsible for previous results? *British Journal of Developmental Psychology*, 1, 289-97.

Simon, T. and Smith P. K. (1985). Play and problem-solving: A paradigm questioned. *Merrill-Palmer Quarterly*, 31, 265-277.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology*, 25, 231-280.

Sloman, S.A. (1996). The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin*, 119, 3-22.

Sloman, S.A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology,* 35, 1-33.

Sloman, S. A. (1999). Cognitive architecture. In R. A. Wilson & F. C. Keil (Eds.), *The MIT Encyclopedia of Cognitive Science*, (pp. 124-126). Cambridge: MIT Press.

Sloman, S.A. (2002). Two Systems Reasoning. In T Gilovich, D. Griffin and D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, New York.

Sloman, S. A. (2014). Two systems of reasoning, an update. In J. Sherman, B. Gawronski, and Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 69-79). Guilford Press.

Smith, E.R. and DeCoster, J. (2000). Dual-Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems. *Personality and Social Psychology Review* 4, 108-131.

Smith, E. E., & Osherson, D. N. (1989). Similarity and decision making. *Similarity and Analogical Reasoning*, 60-75.

Smith, M. (2004). *Ethics and the a priori: Selected essays on moral psychology and meta-ethics*. Cambridge University Press.

Smith, M. (2008). The truth about internalism. In Sinnott-Armstrong, W. (Eds.), *Moral psychology volume 3: The neuroscience of morality: Emotion, brain disorders, and development* (pp. 207-215). New York: Oxford University Press.

Smith, P.K. (1995). Play, ethology, and education: A personal account. In Pellegrini, A. D. (Ed.), *The future of play theory: A multidisiplinary inquiry into the contributions of Brian Sutton-Smith.* SUNY Press.

Smith, P. K. and Dutton, S. (1979). Play and training on direct and innovative problem-solving. *Child Development,* 50, 830-836.

Sober, E. (1999). Testability. *Proceedings and Addresses of the American Philosophical Association,* 47-76.

Sperber, D. and Mercier, H. (2009). Reasoning as a social competence. In J. Elster and H. Landemore (eds.), *Collective Wisdom*, MIT Press.

Stanovich, K. (1999). *Who is rational? Studies of individual differences in reasoning.* Mahwah, NJ: Lawrence Erlbaum.

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

Stanovich, K. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In Evans, J. St. B. T. and Frankish, K. (Eds.), *In Two Minds and Beyond* (pp. 55-87). New York: Oxford University Press.

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

Stich, S. P. (1990). *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation*. The MIT Press.

Tardif, T., Gelman, S. A., Fu, X., and Zhu, L. (2012). Acquisition of generic noun phrases in Chinese: learning about lions without an '-s'. *Journal of Child Language, 39,* 130-161.

Tavares, M. C. H., & Tomaz, C. (2002). Working memory in capuchin monkeys (Cebus apella). *Behavioural brain research, 131*(1), 131-137.

Tversky, A., and Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review,* 90, 293-315.

Tverky, A. and Kahneman, D . (1986). Rational choice and the framing of decisions. *Journal of Business, 59,* 251-278.

Port, R., and van Gelder, T. (Eds.). (1995). *Mind as motion*. Cambridge: MIT Press.

Wason, P. C. (1966). Reasoning. In B.M. Foss (Ed.), *New horizons in psychology I*. Harmondsworth: Penguin.

van Fraassen, B. C. (1980). *The scientific image*. Oxford University Press.

Vogel, J. (1990). Are there counterexamples to the closure principle? In Roth, M. D. and Ross, G. (Eds.), *Doubting: Contemporary perspectives on skepticism* (pp. 13-29). Dordrecht: Kluwer.

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

Wilson, T., Lisle, D., Schooler, J., Hodges, S., Klaaren, K., and LaFleur, S. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Personality and Social Psychology Bulletin,* 19, 331-339.

Wright, Larry. (1976). *Teleological explanations: An etiological analysis of goals and functions*. Berkeley and Los Angeles: University of California Press.

Xu, F. and Pinker, S. (1995). Weird past tense forms. *Journal of Child Language* 22, 531–556.