*To appear in In Two Minds: Dual Process and Beyond (eds.) Frankish K. & J. Evans, OUP.*

# The Magical Number Two, Plus or Minus: Dual Process Theory as a Theory of Cognitive Kinds

**Richard Samuels**

## *Introduction*

A central explanatory objective of much cognitive science is to provide an account of our *mental architecture*: to characterize the various systems and structures from which the human mind is composed. In this regard dual-process theories are unremarkable. For they too are in the business of providing (at least partial) answers to questions about what cognitive systems there are, and what properties they possess. Indeed dual-process theorizing has proven extraordinarily fruitful, and resulted in the development of a broad array of hypotheses about the systems responsible for such apparently disparate phenomena as learning (Reber, 1993), deductive reasoning (Evans 2002), probabilistic judgment (Barbey & Sloman, 2007), decision-making (Kahneman & Frederick, 2002) and social cognition (Chaiken &Trope, 1999; Leiberman, 2003).

Whilst these hypotheses are strikingly similar, they also differ in crucial respects. Not only do they concern different domains of cognition, they also attribute different properties to the cognitive processes and systems they seek to explain. This raises some quite fundamental issues about dual process theories in general. Most obviously, it raises the issue of how best to think about the relationship between the various hypotheses currently on offer. There are two broad options. According to the first, extant hypotheses are merely related by family resemblance –e.g. in positing two processes or systems with broadly similar characteristics. According to the second, the various extant hypotheses are instances of some generic dual process account of cognition. No doubt, there are dual-process theorists who would settle for the first of these options. Yet the second is clearly the more intriguing, and moreover one that's more-or-less explicit in the research of some prominent dual-process theorists (e.g. Evans 2008; Stanovich, 2004). In what follows, it is this second option –the prospect of a generic dual-process account—that I focus on.

Specifically, I consider the question of whether there is some general, distinctive and plausible dual-process thesis about cognition. I argue that there is such an account –what I call the *type* (or cognitive kinds) version of dual process theory.

Here's how I'll proceed. In section 1, I identify a substantive and interesting class of hypotheses worthy of the 'dual process' label and contrast them with some other rather more banal claims. Then in section 2, I distinguish between two quite different versions of the hypothesis: one on which dual-process theory is construed as a thesis about cognitive tokens (or particulars), the other on which it is a thesis about cognitive types (or kinds). These two versions of dual process theory differ considerably in plausibility. Or so I maintain. In section 3, I sketch the reasons for rejecting the cognitive tokens version of dual-process theory; and in section 4, I highlight the virtues of the cognitive types version. Finally, I conclude in Section 5 by arguing that given a clear understanding of the cognitive types proposal we can deflect the main general objections to dual process theorizing.

## *1. How to Understand Dual-Process Theories*

The dual-process proposal is, I maintain, a bold and substantive one. But to see this we need to get clearer on its commitments and distinguish it from some superficially similar ones. Let me start, then, by saying what the dual-process proposal is not.

### *1.1 Dual-process accounts claim far more than that there is <u>some</u> bipartite division of cognitive processes.*

Though dual-process theories may sometimes appear to say no more than this, the appearance is (fortunately) misleading. If this were the view, it would be a banal and uninteresting one. There are always *lots* of ways of dividing up processes. This is true not only of cognitive process, but processes quite generally. (Exercise: Think of all the ways of subdividing a process, such as making chocolate.) How many processes one acknowledges will depend on the *granularity* of the distinctions one draws; and it is very easy to find *some* bipartite distinction or other for almost any (kind of) process.

One respect in which dual-process theories go beyond the above banal claim is in characterizing mental processes in terms of a menu of distinctions that are widely

regarded as important to understanding cognition, such as the distinctions between associative and rule-based processes, and between conscious and unconscious ones. As we will see later, dual-process theorists vary in the distinctions they invoke. But the following table contains the most common ones: what I'll call the *Standard Menu* of distinctions. (For the moment, ignore the "S1" and "S2" labels. I'll come back to these very soon.)

| **S1** | **S2** |
|---|---|
| Associative | Rule-based |
| Heuristic | Analytic |
| Parallel | Serial |
| Automatic | Controlled |
| Unconscious | Conscious |
| Low demands on cognitive capacity | High demands on cognitive capacity |
| Relatively fast | Relatively slow |
| Contextualized | Decontexualized |
| Evolutionarily Old | Evolutionarily new |
| Conserved across species | Unique to humans |

Table 1: *Standard Menu.*
Distinctions between cognitive processes

*1.2 Single and Covariant Factors*

Dual-process theories do not claim merely that cognitive processes can be divided in terms of single dichotomies on the Standard Menu. No doubt some cognitive processes are more automatic than others, some faster than others, and so on. No doubt, some of these distinctions are more interesting than others; and some of them harder to specify with precision. (Further: some of the pairs of terms on the Standard Menu almost certainly correspond to more than just one distinction.) But it should be relatively uncontentious that *some* such distinctions can be drawn. So, a theory that claimed merely that such distinctions can be drawn –that there are, for example, relatively fast and

relatively slow processes— would be a rather uninteresting one. But fortunately, the dual process theorist's position is far bolder than this. What they claim is that such distinctions *line-up*: that processes which exhibit one property from a column typically, though not invariably, possess the others. According to this view, then, the properties listed under 'S1' and the properties listed under 'S2' form *clusters* of *co-varying* properties. Thus a central and far from banal commitment of dual-process theories is that cognitive processes –either generally or within some domain, such as reasoning—can be divided in two: those that possess the S1-property cluster and those that possess the S2-property cluster.

*1.3. So What? From Property Clusters to Cognitive Mechanisms*

Why should it matter whether S1-properties and S2-properties tend to cluster? One important reason is that it provides the basis for a familiar (and wholly respectable) kind of argument for the existence of a division between cognitive *systems*. Thus the existence of clusters bears on what I earlier identified as perhaps the central project of cognitive science: the task of characterizing the mind's architecture.

Let me spell out the argumentative strategy. By assumption, S1-properties and S2-properties form covarying clusters. But the members of each cluster are not logically dependent on each other in the way that, for example, being a bachelor is logically dependent on being unmarried. Nor is the clustering plausibly viewed as a product of mere coincidence. So, *why* do the properties form clusters? Some kind of explanation is required; and one very common, general strategy is to posit an underlying suite of mechanisms to explain covariation. This is a specific instance of what philosophers call an inference to the best explanation. In the present case, dual-process theorists propose a division between cognitive mechanisms that subserve processes exhibiting the S1 cluster and those that subserve S2 exhibiting processes. Moreover, to the extent that the explanation is the best one, we have reason to accept it. Thus according to dual-process theorists, not only do cognitive processes exhibit disjoint clusters of properties, but this putative fact also provides reason to posit a bipartite division between cognitive *mechanisms*.

Here's a philosophically more loaded way to put the point. The existence of cognitive processes that exhibit distinct property clusters enables us to identify a bipartite division in the *natural kinds* that underwrite cognition. The expression 'natural kind' has of course been explicated in many different ways. But the most popular, and to my mind most plausible, view to have emerged from recent philosophy of science is that natural kinds are *homeostatic property clusters* (Boyd, 1991). Roughly put, according to this view, a kind is natural if:

i. It is associated with a range of characteristics or symptoms, which tend to be co-instantiated by instances of the kind, but need not be genuine necessary conditions for membership.
ii. There is some set of underlying causal mechanisms and constraints –a 'causal essence', if you will— whose operation explains the co-instantiation of these various symptoms.
iii. To the extent that there is any real definition of what it is for something to be a member of the kind, it is not symptoms but causal essence that define membership.

Consider an illness such as influenza. Influenza is, on the homeostatic cluster view, a plausible candidate for natural kind status. First, it is associated with a range of characteristic symptoms –coughing, elevated body temperature, and so on—even though these symptoms do not *define* what it is to have flu. Second, there is a causal mechanism –roughly, the presence of the flu virus— whose operation explains the occurrence of the symptoms. Finally, to the extent that influenza has a definition, it is the presence of the virus –or better, the presence of the virus producing some symptoms— but not the symptoms as such, that make it the case that one has flu.

The proposal is intended to capture paradigmatic examples of natural kinds, such as water and influenza. But it also comports quite well with the project of cognitive science where researchers are often in the business of characterizing the mechanisms that subserve the various cognitive processes on which our behavior depends. More specifically, it comports well with the explanatory objectives of dual process theorists since precisely what they do is posit underlying causal mechanisms in order to explain the existence of processes that exhibit distinct property clusters.

***2. Two versions of dual-process theory: Cognitive tokens and cognitive types***

So far we have seen that dual process theorists endorse the following pair of claims:

> Dual-Cluster Thesis: Cognitive processes tend to exhibit either the S1 or S2 property clusters.
>
> Dual-Systems Thesis: There is a division in our cognitive architecture –a division between cognitive systems— that explains this clustering effect.[1]

From hereon, I will assume that any version of dual-process theory worthy of the name is committed to these generic claims. Yet there are two rather different ways of developing this generic formulation of dual-process theory: what I'll call the token and type theses.

To a first approximation, what the *Token Thesis* maintains is that each mind contains two *particular* cognitive mechanisms or systems. The first, sometimes called *System 1* –though also variously and more expressively, the 'heuristic', 'implicit' or 'associative' system— subserves those cognitive processes that tend to exhibit the S1-property cluster. The second mechanism, often called *System 2* –though also variously referred to as the 'analytic', 'explicit' or 'rule-based' system— subserves those cognitive processes that tend to exhibit the S2-property cluster. On this view, then, each human mind exhibits a fundamental, bipartite division into particular systems.

According to the second, *Type Thesis*, each mind is comprised of two types or *kinds* of cognitive system. Systems of the first kind –type-1 systems— subserve processes that tend to exhibit the S1 cluster. Systems of the second kind –type-2 systems— subserve processes that tend to exhibit the S2 cluster. Such a view is logically weaker than the Token Thesis: the Token Thesis implies the Type Thesis but not vice versa. Even so, the Type Thesis still makes a bold and substantive claim about our minds, namely that each exhibits a fundamental, bipartite division into kinds or types of cognitive system. If the view were correct, then it would identify a central division in the

---

[1] Note: The first claim could be adopted without endorsing the second; but canonically they go together. Further, it is the endorsement of the latter that gives dual process theory much of its import and substance.

systems which comprise our minds. The question I now wish to consider is this: Is either version of dual process theory plausible?

## *3. The Token Thesis: Dual-process theory as a theory of cognitive tokens*

Let's start with the Token Thesis. Though the proposal comes in a variety of forms, I maintain that none are attractive. Indeed, it's unclear how to formulate the thesis so as to avoid rendering it either highly implausible or else trivial.

### *3.1. Refining the Token Thesis*

One initial, though obviously unsatisfactory, formulation of the Token Thesis is that our minds contain exactly two cognitive systems: System 1 and System 2. But literally construed this turns dual-process theory into a straw man. One reason is that cognitive scientists quite generally, and dual-process theorists in particular, routinely advocate a strategy of functional decomposition which conceives of relatively complex systems as hierarchically decomposable into relatively simple component subsystems. On any remotely plausible version of this story, there are going to be *far more* than two systems.

The point should be obvious even if we restrict ourselves to, for example, the human vision system. Here, it is overwhelmingly likely that the overall system is organized into subsystems, including, for example, for depth perception, color identification and categorization (Palmer, 1999). Further, it is likely that these subsystems are themselves composed of further subsystems, which further decompose into smaller units, and so on. If anything like this story is correct –and virtually all vision scientists assume it is— then there are obviously going to be more than two systems. Indeed, if we focus on lower levels in the hierarchy of decomposition, we should expect to find *loads* of them –still more if we are permitted to sum mechanisms across all the different levels in the decomposition hierarchy. So, even if we restrict ourselves to a particular region of

cognition, there are likely to be lots of mechanisms; and I presume that no dual-process theorist would seriously claim otherwise.[2]

Of course, none of the above shows that the Token Thesis is false. My aim is merely to eliminate an obvious misreading: one that, if accepted, would turn the thesis into a straw man. But how ought it to be formulated so as to avoid such obvious pitfalls? One possibility is to relativize dual-process theory to a specific level of decomposition. On this suggestion, dual-process theory would not claim there are two systems *tout court* but that there are two systems relative to some –presumably quite abstract— level of decomposition. Yet even at quite abstract levels of decomposition, it's just not plausible that our minds contain only two systems. On any plausible decomposition, there are likely to be a great many systems for a wide range of different mental processes, including perception, memory, reasoning, emotion, language, and no doubt many others. Moreover, it's not plausible to treat all these devices as constituting just two systems. Not, at any rate, unless one is prepared to countenance systems that are wildly heterogeneous in character.

Again, the present observation does not refute the Token Thesis so much as indicate a need for further refinement. And the obvious addition is to restrict the domain to which one's theory is supposed to apply. In doing so, one ceases to treat dual-process theory as a view of cognition in general, and instead reformulates it as a thesis about some specific region of cognition. *Which* region? Clearly, there are many possibilities. Perhaps the most common –and the one most relevant for the purposes of this volume— would be to restrict the claim to the domain of *reasoning*. On such a proposal, what dual process theories (of reasoning) claim is that relative to some appropriate –presumably quite abstract— level of decomposition, there are exactly two *reasoning* systems.

---

[2] This is not, of course, to say that visual processing doesn't also divide up in ways that are conducive to some version of the dual-process theory. So, for example, Milner & Goodale (1995) suggest a dichotomous division within vision –roughly put, a distinction between a system for fast action in the world and another for object recognition—that seems broadly consonant with dual-process approaches to cognition.

*3.2. Two reasoning systems?*

So, the present suggestion is that there are two specific reasoning systems, one whose processes exhibit the S1 cluster, the other whose processes exhibit the S2 cluster. On the face of it, this appears to be a plausible Token Thesis; and moreover, one that some prominent dual-process theorists advocate. So, for example, Steven Sloman appears to defend such a view in his widely cited "The Empirical Case for Two Systems of Reasoning" (Sloman, 1996).

Yet the proposal is hardly free from difficulty. The claim that there are two reasoning systems is not supposed to be mere stipulation, but a substantive and plausible empirical hypothesis. For this to be so, however, there needs to be some way of distinguishing reasoning –or a reasoning system— from the rest of cognition so that there are plausibly just two reasoning systems. The problem is that it is far from clear that this can be done. Consider what appear to be the main options:

*Suggestion 1. Any inferential device is a reasoning system.* Problem: This will not do for the purposes of the Token Thesis since it includes too much. Inference of some sort appears to be a pervasive feature of much cognition, including perception and action guidance. Thus on this construal of reasoning systems, there are going to be too many – i.e. more than two.

*Suggestion 2. Any system that subserves conscious deliberative inference is a reasoning system.* Problem: Though commonplace amongst philosophers and sometimes adopted by psychologists (e.g. Haidt 2001), this characterization of reasoning will not do because it obviously excludes too much. In particular, on such a view, System 1 itself would not count as a reasoning system since it is supposed to subserve automatic and largely unconscious processes.

*Suggestion 3: Any device involved in paradigmatic reasoning tasks is a reasoning system.* Problem: Again, this will not do because it includes too much. A paradigmatic reasoning task –such as a deductive reasoning problem or probabilistic judgment— draws on many psychological capacities, including perception, motor control, language and perhaps much more. In which case, on this view of reasoning systems, there will be too many.

*Suggestion 4: Reasoning systems are to be identified with so-called 'central' systems.* Problem: This is a fudge. Though the distinction between central systems – roughly, those responsible for 'higher' cognition— and non-central systems may be useful for heuristic purposes, it's hard to draw with any precision. Sometimes the distinction is characterized in terms of those systems that subserve reasoning and those that do not (Samuels, 1998). But this clearly won't do for present purposes since it is reasoning that we seek to characterize. Another approach is to characterize central systems as those that deploy *conceptual* representations. Yet this tells us little in the absence a clear distinction between conceptual and non-conceptual representation: a distinction that is notoriously hard to draw.

*Suggestion 5: "If you've got to ask, you're never going to know."* A final suggestion is that we waive the demand for any explicit characterization of reasoning. Instead one insists that, as with many areas of enquiry, our intuitive notions are good enough to give us a reasonable purchase on the phenomena we seek to understand. On this view, our intuitive notion of reasoning –though no doubt rough-and-ready and largely implicit— is good enough for government purposes: sufficiently precise to permit the identification of *bona fide* reasoning systems.

Well maybe so.[3] But if one takes this line, then it's very far from clear there that there are just two reasoning systems. This is because reasoning as we intuitively construe it seems to include a great many kinds of process for which there are good empirical grounds for supposing the existence of distinct cognitive systems. Though there is not the space to discuss the evidence in detail here, plausible candidates include: arithmetic inference (Dehaene 1999; Gelman & Butterworth 2005) probabilistic reasoning, decision-making, planning, spatial reasoning, reasoning about social phenomena, ethical judgment (Greene & Haidt, 2002) reasoning about the minds of others (Leslie, 2004) enumerative induction, and abductive inference. Intuitively, all these things appear to involve kinds of reasoning; and there is little or no reason to suppose that they all depend on just two systems. In short: Though possible, the present suggestion lacks empirical plausibility.

---

[3] Though it should be noted that I have doubts that our intuitive grasp of reasoning is really all that clear.

*3.3 One-way or two-way collapse?*

It's worth stressing that the closing point of the previous section is very widely acknowledged amongst dual-process theorists (Stanovich, 2004; Evans, 2007). But the most common response is to adopt a view on which the Token Thesis is only *half* wrong. On this view, though the Token Thesis is wrong to claim that there is a single System 1, it is right about System 2: there really is just one per mind. As Keith Stanovich has put it:

> I have used terms such as System 1 or heuristic system as if I were talking about a singular system… However, using a term such as heuristic system –which implies a single cognitive system—is really a misnomer. In actuality, the term used should be plural because it refers to a (probably large) *set* of systems in the brain that operate autonomously in response to their own triggering system, and are not under control of the analytic processing system (Stanovich, 2004, p. 37).

In short: The collapse of the Token Thesis only goes in one direction.

Why suppose the collapse is unidirectional: that there is still just one System 2 per head? This is unclear; and of late, some dual-process theorists have suggested the need for additional systems. Stanovich has, for example, tentatively proposed a division within System 2 between algorithmic and reflective systems (Stanovich, this volume). But it seems to me that there are at least two considerations that militate in favor of a more extensive plurality.

*Consideration 1:* The first depends on an observation about the present state of research within the dual-process tradition. As noted earlier, dual-process theorizing has been applied to a wide range of cognitive domains, including learning, deductive reasoning, probabilistic judgment, decision-making, and social cognition. In each domain one finds apparently well-motivated empirical hypotheses that seek to explain the available range of data for that domain. Moreover, these various hypotheses posit systems and processes that bear striking similarities to each other. Yet the processes and mechanisms posited in each domain are *not* characterized in identical fashion. They also *differ* in crucial respects. For instance, they differ in how they characterize what the systems are supposed to do –execute rules, as opposed to exert executive control, for example. Moreover, their processes are not characterized in terms of the very same

properties from the Standard Menu. So, for example, in the domain of social cognition, Leiberman and colleagues have posited a mechanism for controlled social cognition –the C-System— whose processes exhibit much of the S2 cluster but also differ in various respects from the analogous system(s) posited by researchers working on, say, deductive reasoning (Lieberman, 2003). Now one possibility is that different researchers are groping towards the identification of a single System 2 for reasoning that is operative across all these various domains. But an alternative is that researchers in different domains may well be identifying different mechanisms that subserve processes of the same general type –i.e. those exhibiting the S2 cluster. Further, this is, I maintain, a *preferable* view. For unlike the single system view, it does not demand on general grounds that we treat the differences between dual-process hypotheses as steeped in error. Instead it permits us to take seriously not only the similarities but also the differences between these more specific –and empirically more motivated— hypotheses. And this, it seems to me, is a *prima facie* reason for not adopting the single System 2 position.

*Consideration 2:* A second *prima facie* reason for not adopting the single System 2 position is that the processes allegedly subserved by this system appear to exhibit precisely the sorts of *heterogeneity* that should lead us to suspect that they depend on multiple systems. On many versions of the dual-process theory, there is a wide variety of cognitive processes –including, planning, decision-making, deductive inference, the construction of self narrative, causal-mechanical reasoning, and the production of explanation— that are, at least some of the time, subserved by System 2 (Stanovich, 2004). Given a commitment to a single System 2, this inclusiveness seems wholly appropriate. For by assumption, System 2 is supposed to subserve all S2-exhibiting processes, which routinely include instances of the above sorts of processes. For example, there are many occasions when our planning and deductive inference are slow, controlled, conscious, serial, and cognitively demanding. *Mutatis mutandis* for explanation, causal reasoning, self-narrative and the rest. Even so, I maintain that it is not at all plausible to view all S2-exhibiting instances of planning, deductive inference, causal explanation and the like, as the products of a single cognitive mechanism. On the contrary, these processes differ in precisely those respects that should lead us to conclude

that they involve different cognitive systems. Specifically, they differ in ways that are typically taken as relevant to the *individuation* of cognitive mechanisms.

First, they differ considerably in *functional* respects. Consider planning, for example. In contrast to, say, deductive inference or causal explanation, planning is centrally concerned with the guidance of action –with identifying sequences of behaviors that collectively facilitate the attainment of our goals. Moreover (and presumably because of this) planning involves a quite different mapping from inputs to outputs than those found in deductive inference or causal reasoning. Most obviously, a planning process takes both beliefs and goals (or desires) as input and generates plans (or intentions) as output, whereas other sorts of process –deductive inference or causal explanation, for example— do not.

Second, and equally importantly, the *computational* demands of the various processes appear to be quite different. At any rate, the most plausible computational models to have emerged from cognitive science and artificial intelligence –for planning, deductive reasoning, causal explanation, and so on— vary considerably both in the sorts of computations that are taken to be relevant and the representational formats that are used (Russell & Norvig, 2003). But if this is so, then to the extent that we take attempts at computational modeling seriously, we should suppose such processes differ in important computational respects.

Now, of course, it is quite possible for a single system to underwrite multiple processes. But the present point is not merely that various S2-exhibiting processes differ, but that they differ in *precisely* those respects that militate in favor of multiple systems. For psychological mechanisms are, by very widespread consensus, individuated by their functional and computational properties. And if this is so, then the assumption that, for example, planning and causal explanation depend on the very same system is implausible. It would thus seem that there is *prima facie* reason for skepticism about the claim that all instances of S2-exhibiting processes are subserved by a single System 2.

**4. The Type Thesis: Dual-process theory as a theory of cognitive types**

In the preceding section I argued that the Token Thesis is implausible. There is unlikely to be just one System 1 or just one System 2. What of the Type Thesis? How

plausible is the claim that our minds are composed of two different *kinds* of cognitive mechanism –type-1 systems that exhibit the S1-cluster and type-2 systems that exhibit the S2-cluster?

The short answer is that it's simply too early to tell. Our understanding of cognitive architecture is at best radically incomplete; and nowhere more so than when it comes to the study of such 'higher' cognitive capacities as reasoning. Nonetheless, I maintain that the Type Thesis remains a genuine contender –a plausible thesis worthy of serious consideration. My reasons for this claim are straightforward. Dual process theorizing is worthy of serious consideration because it earns its *explanatory keep*. In particular, it has generated a wide array of fruitful explanations for many phenomena. To select some examples more-or-less at random:

- *Belief Bias:* In the psychology of deductive reasoning, dual process theorizing constitutes perhaps the standard approach to explaining belief bias. Roughly put, it explains our tendency to assess the validity of arguments quite differently depending on the believability of their conclusions (Evans & Over, 1996).
- *Interpersonal Variation in Reasoning:* Dual process theories have proven fruitful in explaining interpersonal variation on standard reasoning tasks and, moreover, explaining why performance correlates with such factors as psychometric intelligence and cognitive style (Stanovich, 1999; Stanovich and West, 1997).
- *Probabilistic Judgment:* Dual process theories have been invoked to explain the patterns of data surrounding such phenomena as base-rate neglect and the conjunction fallacy (Barbey & Sloman, 2007)
- *Cross-Cultural Variation:* Dual process accounts have held centre stage in recent attempts to explain cross-cultural differences in cognition, such as those characteristic of Western and South-East Asian subjects (Norenzayan et al. 2002).

Though there is not the space to defend the claim here, I maintain that these and other examples of dual-process theorizing have resulted in plausible empirical hypotheses. But as I have already argued, the Token Thesis is implausible because it seems likely that there are both many system 1's and many system 2's. In which case, if dual process theories are worthy of serious attention, then it would seem that we are committed to the Type Thesis.

Of course, such an argument would the undermined if the Token Thesis were required for the above kinds of explanations to work. But this is not the case. On the contrary, the commitment to just two particular systems is gratuitous. In the case of belief bias, for example, what's required to explain the fact that we're more likely to assess arguments as valid when their conclusions are plausible is not the assumption of *exactly* two (reasoning) systems. What's required is that, when assessing the validity of arguments, two competing systems are involved. But this is wholly consistent with these two systems being two systems amongst many others. Thus not only is the Type Thesis the only plausible available version of dual-process theory, but it would also seem to suffice for the explanatory purposes that make dual process approaches attractive in the first place.

## 5. Some Challenges

I have suggested that the Type Thesis is a plausible one worthy of serious consideration. Nonetheless, the claim is not without its problems. As one might expect, much of the action concerns the details of specific empirical hypotheses and the nature of the evidence that has been invoked in their defense. But there are also some quite general concerns that have been expressed by both opponents and friends of dual process theorizing (Evans, 2006). I propose to close by considering three such problems. I maintain that whilst each calls for substantial further research, none pose insurmountable problems for the type version of dual process theory.

### *5.1 Problem 1: The Specification Problem*

In characterizing dual-process theory, I have so far alluded vaguely to two different property clusters, S1 and S2. But this is an oversimplification. Though there is substantial overlap between different formulations of dual-process theory, the characteristics invoked by different versions are not identical. In view of this, an obvious question arises: What precise characterizations of the S1 and S2 property clusters ought we to adopt?

In fact, there are two closely related issues here. The first, 'conceptual problem', concerns how to characterize in a precise and appropriate manner the various dichotomies

invoked by dual process theorists. This problem arises, in large measure, because the distinctions have been characterized in a variety of non-equivalent ways. So, for example, there exist a variety of formulations of the distinction between controlled and automatic processes (Moors and De Houwer, 2006). A fully developed version of dual-process theory should presumably address this worry.

The second, 'inclusion problem', concerns which distinctions should figure in our characterization of the two property clusters. Should we invoke, for example, a distinction between parallel and serial processes as Sloman does, or should we resist this suggestion, as Evans does? *Mutatis mutandis* for conscious versus unconscious, controlled versus automatic, evolutionarily ancient versus novel, and so on. Again, a fully developed version of dual-process theory should presumably address this worry.

To what extent should the above be a cause for concern? The answer is, I think, not much. There are two reasons for this. First, on the assumption that the Type Thesis is correct, the lack of clarity about property clusters is unsurprising. This is because it is quite possible for distinct, individual mechanisms to differ in *some* respects even if they are members of the same broad class of mechanisms.[4] Second, even waiving this consideration, the fact that the property clusters require revision and clarification is hardly reason to reject dual process approaches. On the contrary, the need to refine general hypotheses in this way is just a commonplace feature of many scientific enterprises. That is, in the course of time, general hypotheses get refined and elaborated. At this time, there is no reason to suppose that dual-process theorizing is any different in this regard. In effect the present problems are just that: problems for future research, and not substantive objections.

*5.2 Problem 2: The Crossover Problem*

There is, however, a second, closely related worry about dual-process theories that on the face of it poses a more serious challenge. According to this worry, the problem is not merely that dual-process accounts are underspecified or that some extant formulations are incorrect. Rather the concern is that no interesting type-1/type-2

[4] Thanks to Chris Viger for this point.

distinction can be drawn because the characteristics exhibited by cognitive processes are not amenable to a clean bipartite division into two property clusters. On this view, the worry is that there are many 'crossover' processes that exhibit a combination of both S1 *and* S2 properties. Consider the property of evolutionary recency, for example. According to many standard formulations of dual process theory, processes that exhibit the S1 property cluster are evolutionarily ancient, whilst those that exhibit the S2 cluster are paradigmatically recent and unique to humans. Yet there appear to be evolutionary recent processes that possess many S1 properties. So, for example, judgments of numerical magnitude involving familiar, conventional numeral systems –such as Arabic numerals— possess many S1 properties. For instance, they are relatively fast (≈200ms); and appear to be automatic in that they exhibit Stroop-like effects –so-called 'number Stroop'— where we automatically access our sense of the numbers designated by numerals and order them by number size, even when number size is irrelevant to the task at hand. (Girelli et al. 2001). Nevertheless, the process so characterized is presumably an evolutionarily recent one since it depends on cultural innovations –conventional numeral systems—which are themselves very recent. Thus the process involved in judgments of numerical magnitude would appear to exhibit both S1 and S2 properties.

How serious is this Crossover Problem? In particular, should the existence of crossovers lead us to reject dual-process theory? The answer is, I maintain, that it poses no serious problem, so long as crossovers are not too numerous or too extreme. One possibility is that they merely indicate the need to drop some notions –e.g. evolutionary recency—from our characterization of the property clusters. And as already noted, such modest revisions are both commonplace in science and provide no serious grounds for rejecting the Type Thesis as such.

A second, and more interesting possibility is that crossovers can be tolerated without any modification whatsoever to the assumed property clusters. As I noted in section 1, dual-process theorists are most plausibly construed as being in the business of identifying some of the cognitive natural kinds that underlie our mental processes. But on such a view, the members of S1 and S2 are not necessary conditions on the activity of their associated systems. That is, the presence and activity of some system 1 (or system 2) does not require that all members of its associated property cluster obtain, any more

than, say, the presence of all the typical symptoms of influenza is a necessary condition for having flu. The point is that, for natural kinds quite generally, the relationship between underlying mechanism and associated characteristics is far weaker than this. In which case, saddling dual-process theorists with the requirement that system activity is always accompanied by all members of its associated property cluster is simply too demanding.

*5.3 The Unity Problem*

The final problem that I discuss here is what I call the *Unity Problem*. Though positing mechanisms is a standard strategy for explaining the existence of property clusters, it does not, by itself, constitute a satisfactory explanation. Rather one needs to specify those features of the proposed mechanisms that account for such clustering effects. In the present case, we need to specify those characteristics of type-1 systems that yield S1-exhibiting processes, and those properties of type-2 systems that yield S2-exhibiting processes. Again, this does not strike me as a serious objection so much as a challenge for future research –one that requires a more detailed account of the systems responsible for type-1 and type-2 processes. Nevertheless it is a significant explanatory challenge that has not, to date, been fully addressed. In what follows I comment briefly on three of the more plausible recent suggestions.

*Proposal 1: Associative and Rule-based Systems.* One general approach to explaining the type-1/type-2 distinction would be to assume the cognitive mechanisms that underlie them operate according to different computational principles. Such a view is suggested, for example, by Steven Sloman's version of dual-process theory.[5] According to this view, the existence of distinct S1 and S2 property clusters can be explained on the assumption that the mind has a hybrid classical/connectionist architecture –one in which type-1 systems are connectionist (associative) mechanisms, and type-2 systems are classical computational (rule-based) ones.

This suggestion is not without its virtues. First, the general form of the proposal is of the right sort to address the Unity Problem. If the style of computation implemented by

[5] I say 'suggested' because it's not at all clear Sloman endorses the suggestion.

type-1 systems were appropriately different from that of type-2 systems, then it would also be unsurprising that the character of their respective processes differed significantly. Second, the solution would be a theoretically elegant and simple one since it would reduce the type-1/type-2 distinction to the distinction between classical and connectionist computational devices. Finally, the idea that cognition depends on both classical and connectionist devices is not without plausibility; and many recent theories of the mind incorporate some version of this suggestion (Anderson, 2003; Sloman, 1996).

For all that, there is little reason to accept the proposed solution to the Unity Problem. First, aside from theoretical elegance, there is currently no reason to suppose that the type-1/type-2 distinction corresponds to the connectionist/classical divide. So, for example, special-purpose learning mechanisms of the sort studied by Gallistel and his collaborators –for dead reckoning, path integration and geometric inference, for example—are often treated by dual-process theorists as paradigmatic examples of type-1 systems (Sloman, 1996; Evans, 2008). But is it far from clear that such mechanisms have a connectionist architecture. Indeed, Gallistel's own account of these mechanisms is militantly classical in character (Gallistel, 1990).

Another problem is that it's very unclear how the classical/connectionist distinction is supposed to explain many of the properties associated with type-1 and type-2 processes. Why, for example, should only classical processes be consciously accessible and controlled? Clearly, neither are essential properties of classical systems. Indeed it is hard to see why classical processes should be any more controlled or accessible to consciousness than connectionist ones. Conversely, why should connectionist processes be more automatic or less cognitively demanding than classical ones? On the face of it, there is no reason to suppose that they should. In short: it is, to put it mildly, obscure why connectionist processes should exhibit the S1 cluster whilst classical ones exhibit the S2 cluster. But absent an explanation of this (putative) fact, presenting the classical/connectionist distinction as a solution to the Unity Problem is mere hand waiving: a gesture at explanation almost wholly lacking in content.

*Proposal 2: Cyclic Realization of Type-2 Processes.* A second possible approach to the Unity Problem, recently suggested by Peter Carruthers and Keith Frankish, is that the difference between the characteristics of type-1 and type-2 processes results from the

fact that system-2 is realized by cycles of system-1 operations (Carruthers, this volume; Frankish, this volume). On this approach, the distinctiveness of type-2 processes is explained by providing an empirically rich account of how different type-1 processes combine to realize type-2 systems. For example, Carruthers sketches an elaborate and intriguing account on which System-2 is realized by complexly interacting type-1 systems for language, perception, practical reasoning and many other things besides.

What are we to make of such proposals? First, it's worth stressing that the general claim that type-2 systems depend on type-1 systems –or at any rate mechanisms very much like them— should be relatively uncontentious amongst cognitive scientists. This is because it is widely assumed that a) complex cognitive systems decompose into simpler mechanisms –ultimately into primitive processors— and b) that cognitive explanations proceed by explaining the behavior of complex systems in terms of the operations and interactions of their parts (Bechtel & Richardson, 1993). Primitive processors are more-or-less by definition automatic, unconscious and cognitively undemanding. This much is required to avoid a range of familiar homunculi regress worries. (For instance, if each system responsible for conscious thought were, in turn, dependent upon the activity of conscious subsystems, then a regress would appear to threaten, where each system is dependent on the activity of some further conscious system.) Further, on the plausible assumption that evolution operates by reusing pre-existing types of structure (Marcus, 2004), primitive processors are likely to be evolutionarily more ancient and more widely conserved across species than the more complex systems they realize.

So, the general idea that type-2 systems are realized by mechanisms very much like type-1 seems plausible. But it is one thing to accept this general claim –what amounts to little more than an endorsement of standard methodology in cognitive science— quite another to provide a detailed response to the Unity Problem. And at this time such suggestions are, at best, highly speculative. Consider Carruthers' view, for example. Though this is not the place for detailed discussion, I do think that it possesses some notable virtues. Most obviously, it makes a good attempt at explaining the clustering of S2 characteristics. (See Carruthers, this volume.) But in order to do so, Carruthers helps himself to a very rich 'boxology' of cognitive mechanisms. And this comes at an obvious cost. Many of the assumptions he makes about our cognitive

systems, their properties and relations to each other concern matters about which little is known, and are consequently highly speculative and contentious. Here's a fairly characteristic example. In developing his view, Carruthers provides a central role for inner speech, and then attempts to explain how inner speech operates:

> Here is how it works. In light of the subject's beliefs and goals, a speech action-schema is formulated by the language production sub-system. While overt action is suppressed, an efference copy of the motor instructions is transformed via an emulator system into an auditory representation of the sounds that would have resulted had the action been carried out. This representation is globally broadcast in the manner of conscious images generally, and is received *inter alia* by the language comprehension sub-system. The latter constructs an interpretation of the utterance in the normal way, and presents that (attached to the sounds) to the various other System 1 inferential systems. Hence, just as is the case with external 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. (Carruthers, this volume)

Now I do not wish to claim that this story is incorrect. (I have no idea.) Nor do I wish to claim that a response to the Unity Problem which incorporates such a story is without merit. On the contrary, as an answer to a 'how possible' question – "How could a system exhibiting the S2 cluster be built from a suit of subsystems?"— such an account may well be useful. My point is merely that such proposals are highly speculative and lack much in the way of empirical confirmation. In short: the proposal buys the ability to explain the properties of type 2 processes at the cost of lacking empirical support.

*Proposal 3: Working Memory.* The final response to the Unity Problem that I will discuss here is that the existence of the S1 and S2 clusters depends crucially on the nature of human working memory (Evan, 2008). According to this proposal, what generates the division between type-1 and type-2 processes is that the latter, but not the former, "require access to a single, capacity limited central working memory resource" (Evans, 2008). As a consequence, type-2 processes possess certain characteristics lacking in type-1 processes. For example, type-2 processes are slow, sequential and capacity limited because they inherit these characteristics from central working memory.

I think that the present suggestion is worthy of very serious consideration. Indeed it is the most plausible response to the Unity Problem that I know of. One virtue is that it

readily explains some core aspects of type-2 processes. So, for example, on the present view, the sequential character of type-2 reasoning turns out to be a direct consequence of the fact that working memory imposes a sequential 'bottleneck' on the processing of information. A second virtue is that the proposal helps explain what might otherwise appear to be puzzling facts about cognition. So, for example, it suggests an explanation of the fact that individual differences in working memory capacity, in reasoning and in psychometric *g* are highly inter-correlated (Stanovich, 1999; Evans 2008). Roughly put, on the present proposal, these variables are highly correlated because they all depend to a significant degree on the operation of a central working memory system whose capacity varies from individual to individual. A third virtue of the present view is that it comports well with the position defended in this paper –that there are many type-2 systems— since it is quite possible for many distinct systems to share central working memory resources. Finally, and in contrast to the proposals discussed earlier, the present response to the Unity Problem requires only relatively austere assumptions about our cognitive architecture –viz. the existence of a sequential, capacity limited working memory— assumptions that are both widespread amongst cognitive psychologists and well supported by empirical research (Baddeley, 2007; Barrett et al., 2004). In short: the working memory proposal seems to be an elegant and empirical well-motivated response to the Unity Problem.

For all that, the present proposal is clearly in need of further development; and in current form, it fails fully to address the Unity Problem. A first and relatively minor worry is that although the proposal does a good job of explaining some features of type-2 processes –e.g. their serial and capacity limited character— it is far less clear how it is supposed to accommodate others. So, for example, it is less clear why type-2 processes should be more accessible to consciousness or more readily under cognitive control. To explain these putative features of type-2 processes, far more would need to be said about the nature of working memory. My point is not that this cannot be done. Indeed there are clearly possible ways to develop the account. My point is merely that such additions to the proposal are required.

A second, and more serious worry about the working memory proposal is that in its present form it really only addresses one half of the Unity Problem. To address the

Unity Problem it is not enough to specify those characteristics of type-2 systems that yield S2-exhibiting processes. One must also explain why type-1 systems yield S1-exhibiting processes. Moreover, to do this, it is not enough merely to make the more-or-less definitional point that type-1 systems tend to yield such processes. What is required is some account of the characteristic properties of type-1 systems that tend to yield processes of this sort. But in its current form the working memory proposal is almost entirely silent on this matter. Instead type-1 systems are merely characterized by exclusion –as those systems that do *not* rely on central working memory. My point is that even if the present proposal is broadly right about type-2 systems, we still require a substantial account of why systems that operate independently of working memory yield processes that exhibit the S1 cluster. Why, for example, are they relatively fast? Why are they not decontextualized? Why are they not under conscious control? And so on. Absent such an explanation, the present proposal is, at best, incomplete.

***Conclusion***

This chapter has been concerned with the issue of whether or not there exists some generic version of dual-process theory that is both interesting and plausible. I started by identifying those characteristics that an interesting version of the theory would need to possess. I then argued that the most plausible view of this sort is one that reconstructs the original System 1/System 2 distinction as a distinction between kinds or *types* of psychological system. Moreover, I argued on grounds of explanatory value that dual-process theorizing quite generally is worth taking seriously and, consequently, that – by virtue of being the most plausible version of the theory—the Type Thesis is worthy of very serious consideration. Of course, this hardly establishes the truth of the view. As we have seen, there are a number of conceptual and empirical problems that need to be addressed. Nonetheless, if the arguments of the last section are correct, then these problems are not so much reasons to reject the Type Thesis as issues that call for further research.

**Bibliography**

Anderson, J. R., Lebiere, C. (2003) "The Newell Test for a theory of cognition" *Behav. Brain Sci* 26 (5): 587-601.

Baddeley, AD. (2007). Working memory, thought and action. Oxford: Oxford University Press.

Barbey AK, Sloman SA. In press. Base-rate respect: From ecological validity to dual processes. *Behav. Brain Sci.*

Barrett, L. F., Tugade, M. M. & Engle, R. W. (2004) Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, *130*, 553-573.

Bechtel, W. and Richardson, R. C. (1993) *Discovering complexity: Decomposition and localization as strategies in scientific research.* Princeton: Princeton University Press.

Boyd, R (1991) Realism, anti-foundationalism and the enthusiasm for natural kinds. *Dialectica* 43, 1-2, 5-29

Chaiken S, Trope Y. 1999. *(Eds.) Dual-process theories in social psychology*. New York: Guildford Press.

Chen S, Chaiken S. 1999. The heuristic-systematic model in its broader context. In *Dual-process theories in social psychology*, ed. Chaiken S, Trope Y, pp. 73-96. New York: The Guildford Press.

Evans JStBT. 2002. Logic and human reasoning: An assessment of the deduction paradigm. *Psychol. Bull.* 128:978-96

Evans JStBT. 2003. In two minds: Dual process accounts of reasoning. *Trends. Cogn. Sci.* 7:454-9

Evans JStBT. 2006. The heuristic-analytic theory of reasoning: extension and evaluation. *Psychon. Bull. Rev.* 13(3):378–95

Evans JStBT. 2008. Dual-Processing Accounts Of Reasoning, Judgment And Social Cognition, *Annual Review of Psychology*

Evans, J. and Over, D. 1996. *Rationality and Reasoning*. East Sussex, U.K.: Psychology Press.

Galistel, C. R. and Gelman, Rochel (2005), "Mathematical Cognition", in Keith J. Holyoak and Robert G. Morrison (eds.), *The Cambridge Handbook of Thinking and Reasoning* (Cambridge: Cambridge University Press): 559-588.

Gelman, Rochel and Butterworth, Brian (2005), "Number and Language: How Are They Related?" *Trends in Cognitive Science* 9(1; January): 6-10.

Girelli, L., Sandrini, M., Butterworth, B., & Cappa, S. (2001). Number Stroop performance in normal aging and Alzheimer's type dementia. Brain & Cognition 46 (1-2), 144-149

Greene, J. and Haidt, J. 2002. "How (and where) Does Moral Judgment Work?" *Trends in Cognitive Science* 6: 517-23.

Haidt J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgement. *Psychol. Rev.* 108(4):814-34

Kahneman D, Frederick S. 2002. Representativeness revisited: Attribute substitution in intuitive judgement. In *Heuristics and biases: The psychology of intuitive*

*judgment*, ed. Gilovich T, Griffin D, Kahneman D, pp. 49-81. Cambridge: Cambridge University Press.

Friedman, O., & Leslie, A.M. (2004b). Mechanisms of belief-desire reasoning: Inhibition and selection. *Psychological Science,* 15, 547–552

Lieberman MD. 2003. Reflective and reflexive judgment processes: A social cognitive neuroscience approach. In *Social judgments: Implicit and explicit processes*, ed. Forgas JP, Williams KR, von Hippel W, pp. 44-67. New York: Cambridge University Press.

Marcus, G. (2004) *The Birth of the Mind: How a Tiny Number of Genes Creates the Complexities of Human Thought* Basic Books

Moors, A. and De Houwer, J (2006) Automaticity: A Theoretical and Conceptual Analysis Psychological Vol. 132, No. 2, 297–326

Norenzayan, A., Smith, E. E., & Kim, B., & Nisbett, R. E. (2002). Cultural preferences for formal versus intuitive reasoning. *Cognitive Science*, 26, 653-684.

Palmer, S. 1999. *Vision Science: Photons to Phenomenology*. MIT Press.

Reber AS. 1993. *Implicit Learning and Tacit Knowledge*. Oxford: Oxford University Press.

Russell, S. and P. Norvig (2003) *Artificial Intelligence, A Modern Approach*, Second Edition, Prentice Hall/Allyn&Bacon

Samuels R. 1998. Evolutionary psychology and the mass modularity hypothesis. *Brit. J. Philos. Sci.* 49:575-602

Sloman SA. 1996. The empirical case for two systems of reasoning. *Psychol. Bull.* 119:3-22

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

Stanovich, K. 1999. *Who is Rational? Studies of Individual Differences in Reasoning*. Mahwah, NJ: Laurence Erlbaum Associtates.

Stanovich KE, West RF. 1998. Cognitive ability and variation in selection task performance. *Think. Reasoning* 4:193-230