Chapter 3

Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory?

Keith E. Stanovich

In a recent book (Stanovich, 2004), I spent a considerable effort trying to work out the implications of dual process theory for the great rationality debate in cognitive science (see Cohen, 1981; Gigerenzer, 1996; Kahneman and Tversky, 1996; Stanovich, 1999; Stein, 1996). In this chapter, I wish to advance that discussion, first by discussing additions and complications to dual-process theory and then by working through the implications of these ideas for our view of human rationality.

Dual-process theory and human goals: Implications for the rationality debate

My previous proposal (Stanovich, 1999, 2004; Stanovich and West, 2000, 2003) was that partitioning the goal structure of humans in terms of dual-process theory would help to explicate the nature of the disputes in the great rationality debate in cognitive science. The proposal was that the goal structures of System 1 and System 2 were different, and that important consequences for human self-fulfillment follow from this fact. The analytic system is more attuned to the person's needs as a coherent organism than is System 1, which is more directly tuned to the ancient reproductive goals of the subpersonal replicators (likewise, it is also the case that System 1 is more likely to contain memes that are nonreflectively acquired; see Blackmore, 1999; Distin, 2005; Stanovich, 2004). In the minority of cases where the outputs of the two systems conflict, people will often be better off if they can accomplish an analytic system override of the System 1-triggered output. Such a system conflict is likely to be signaling a vehicle/replicator goal mismatch and, statistically, such a mismatch is more likely to be resolved in favor of the vehicle (which all of us should want) if the System 1 output is overridden. This is why in cases of response conflict, override is a statistically good bet.

From within this framework, I have previously criticized some work in evolutionary psychology and adaptive modeling for implicitly undervaluing instrumental rationality by defending non-normative responses made by many subjects in reasoning experiments. Evolutionarily adaptive behavior is not the same as rational behavior.

Evolutionary psychologists obscure this by sometimes implying that if a behavior is adaptive it is rational. Such a conflation represents a fundamental error of much import for human affairs. Definitions of rationality must be kept consistent with the entity whose optimization is at issue. In order to maintain this consistency, the different 'interests' of the replicators and the vehicle must be explicitly recognized. I think a conflation of these interests is at the heart of the disputes between researchers working in the heuristics and biases tradition and their critics in the evolutionary psychology camp.

My research group has shown that while the response that is consistent with many evolutionary analyses (optimal foraging and so forth) is the modal response on many heuristics and biases tasks, the most cognitively able subjects give the response that is instrumentally rational (Kokis et al., 2002; Stanovich and West, 1998a, 1998b, 1998c, 1998d, 1999; West and Stanovich, 2003; see also De Neys, 2006a, 2006b). Our interpretation of this data pattern was that the evolutionary psychologists are probably correct that most System 1 responses are evolutionarily adaptive. Nevertheless, their evolutionary interpretations do not impeach the position of the heuristics and biases researchers that the alternative response given by the minority of subjects is rational at the level of the individual. Subjects of higher analytic intelligence are simply more prone to override System 1 in order to produce responses that are epistemically and instrumentally rational. This rapprochement between the two camps that West and I have championed has also by advocated in several papers by Samuels and Stich (Samuels and Stich, 2004; Samuels et al., 2002; Samuels et al., 1999) who have argued for a similar synthesis (see also, Evans, 2007). Indeed, such a synthesis could be said to be implicit within the early writings of the original heuristics and biases researchers themselves (Kahneman and Frederick, 2002; Kahneman and Tversky, 1982a, 1996; Tversky and Kahneman, 1974, 1983). As Kahneman (2000) notes, 'Tversky and I always thought of the heuristics and biases approach as a two-process theory' (p.682).

Complicating the generic dual-process model

The main purpose of this chapter, though, is to add some complications to the dual-process view articulated in Stanovich (2004). First to a complication that I believe should generate little controversy. Evans (2006a, this volume) and Stanovich (2004) have both argued that although many theorists use terms such as System 1 or heuristic system as if they were talking about a singular system, this is really a misnomer (see also Carruthers, 2006). In actuality, the term used should be plural because it refers to a *set* of systems in the brain that operate autonomously in response to their own triggering stimuli, and are not under the control of the analytic processing system. I thus have suggested the acronym TASS (standing for The Autonomous Set of Systems) to describe what is in actuality a heterogeneous set.[1]

---

[1] Evans (this volume) revives the type 1/type 2 process terminology of Wason and Evans (1975) and I am largely in sympathy with his suggestion. As with the TASS terminology, Evans' (this volume) usage allows that there may be many different type 1 processes.

For example, many TASS processes would be considered to be modular, as that construct has been discussed by evolutionary psychologists and other cognitive scientists, but TASS is not limited to modular subprocesses that meet all of the classic Fodorian criteria. Along with the Darwinian mind of quasi-modules discussed by the evolutionary psychologists, TASS contains domain *general* processes of unconscious implicit learning and conditioning. Also, TASS contains many rules, stimulus discriminations, and decision-making principles that have been practiced to automaticity (e.g. Shiffrin and Schneider, 1977). And finally, processes of behavioral regulation by the emotions are also in TASS (on the types of processes in TASS, see Brase, 2004; Carruthers, 2002; Cosmides and Tooby, 1992; Evans, 2003, this volume; Sperber, 1994). Thus, TASS processes are conjoined in this category on the basis of autonomy, not modularity—specifically TASS processes respond automatically to triggering stimuli; their execution is not dependent upon input from, nor is it under the control of, the analytic processing system (System 2); and finally TASS can sometimes execute and provide outputs that are in conflict with the results of a simultaneous computation being carried out by System 2.

Theoretically, this complication to dual-process models serves to remind us that learned information is in TASS as well as modules that are the result of evolutionary adaptation. This learned information can be just as much a threat to rational behavior—that is, just as in need of override by System 2—as are evolutionary modules that fire inappropriately in a modern environment. Rules learned to automaticity can be overgeneralized—they can autonomously trigger behavior when the situation is an exception to the class of events they are meant to cover (Arkes and Ayton, 1999; Hsee and Hastie, 2006).

The next complication I wish to introduce concerns the conceptualization of System 2 and it is of perhaps more theoretical import. I will argue that System 2 needs to be understood in terms of two levels of processing—the algorithmic level and the reflective level. We can see this if we consider the logic of TASS override. TASS will implement its short-leashed goals unless overridden by the algorithmic mechanisms implementing the long-leash goals of the analytic system. But override itself is initiated by higher level control. That is, the algorithmic level of the analytic system is conceptualized as subordinate to the higher-level goal states and epistemic thinking dispositions, some of which have been studied empirically (e.g. Cacioppo et al., 1996; Stanovich and West, 1997, 2007). These goal states and epistemic dispositions exist at what might be termed the reflective level of processing—a level containing control states that regulate behavior at a high level of generality. Such high-level goal states are common in the intelligent agents built by artificial intelligence researchers (Franklin, 1995; Pollock, 1995; A. Sloman, 1993; A. Sloman and Chrisley, 2003). My attempt to differentiate System 2 into the two levels of processing was the reason for the provocative title of this chapter which was meant to raise the question of how seriously we should take a tripartite model. In Figure 3.1, I have presented the tripartite proposal in a simple form. In the spirit of Dennett's (1996) book *Kinds of Minds*, I have labeled the traditional TASS (or System 1) as the autonomous mind, the algorithmic level of System 2 the algorithmic mind, and the reflective level of System 2 the reflective mind.

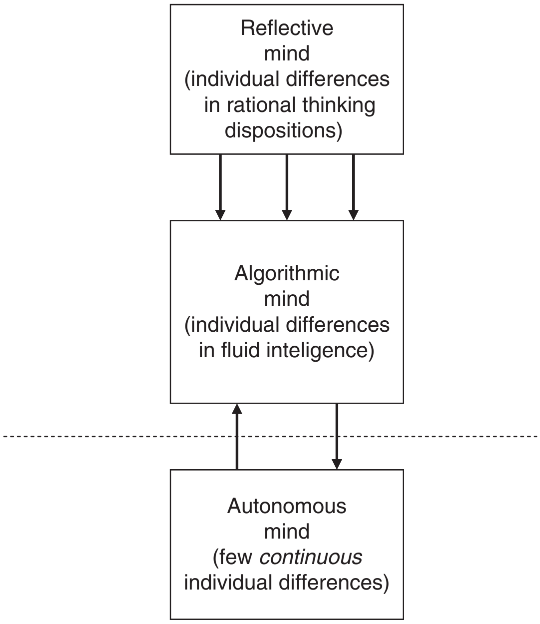


Fig. 3.1 Individual differences in the tripartite structure.

Differentiating the algorithmic and the reflective mind

In fact, artificial intelligence researchers take the possibility of a tripartite structure quite seriously (see A. Sloman and Chrisley, 2003; Samuels, 2005). Are there any other reasons, aside from the precedent in AI research, for accepting this alternative structure? First, it must be acknowledged that there is no way that the distinctions between the algorithmic and reflective mind will cleave as nicely as those that have traditionally differentiated System 1 and 2 (the dashed line in Figure 3.1 signals this) because the algorithmic and reflective mind will both share properties (capacity-limited serial processing for instance) that differentiate them from the autonomous mind.

Nonetheless, there are some reasons for giving the algorithmic/reflective distinction some consideration. My research group has found that individual differences in some very important critical thinking pitfalls such as the tendency toward my side thinking and the tendency toward one-sided thinking are relatively independent of intelligence (Stanovich and West, 2007, in press). We take this to indicate that the critical thinking skills necessary to avoid my side bias and one-side bias are instantiated at the reflective level of the mind as opposed to the algorithmic level. Second, across a variety of tasks from the heuristics and biases literature, it has consistently found that rational thinking dispositions will predict variance in these tasks after the effects of general intelligence have been controlled (Bruine de Bruin et al., 2007; Klaczynski and Lavallee, 2005; Kokis et al., 2002; Parker and Fischhoff, 2005; Stanovich and West, 1997, 1998c, 2000; Toplak and Stanovich, 2002, 2003). Thinking disposition measures are telling us about the individual's goals and epistemic values—and they are indexing broad tendencies of pragmatic and epistemic self-regulation at the intentional level of analysis. The empirical studies cited indicate that these different types of cognitive predictors are tapping separable variance, and the reason that this is to be expected is because cognitive capacity measures such as intelligence and thinking dispositions map on to different levels of analysis in cognitive theory.

Figure 3.1 reflects the theoretical conjecture (Stanovich, 2002, in press a). It is proposed that variation in fluid intelligence (see Carroll, 1993) largely indexes individual differences in the efficiency of processing of the algorithmic mind. In contrast, thinking dispositions index individual differences at the intentional level—that is, in the reflective mind. In the empirical studies cited above, the rational thinking dispositions examined have encompassed assessments of epistemic regulation such as actively open-minded thinking and dogmatism (Stanovich and West, 1997, 2007), assessments of response regulation such as the Matching Familiar Figures Test (Kagan et al., 1964), and assessments of cognitive regulation such as need for cognition (Cacioppo et al., 1996).

The proposal is thus that just as System 1 has been pluralized into TASS, we might now need to recognize to aspects of System 2, the reflective and algorithmic. One reason for endorsing a tripartite structure is that breakdowns in cognitive functioning in the three kinds of minds manifest very differently. For example, disruptions in algorithmic-level functioning are apparent in general impairments in intellectual ability of the type that cause mental retardation (Anderson, 1998). And these disruptions vary quite continuously. In contrast, continuous individual differences in the autonomous mind are few. The individual differences that do exist largely reflect damage to cognitive modules that result in very discontinuous cognitive dysfunction such as autism or the agnosias and alexias. Importantly, Bermudez (2001; see also Murphy and Stich, 2000) notes that they are traditionally explained by recourse to subpersonal functions (see Davies, 2000, for a discussion of personal and subpersonal constructs). In complete contrast are many psychiatric disorders (particularly those such as delusions) which implicate intentional-level functioning (that is, functioning in what I here call the reflective mind). Bermudez (2001) argues that the 'impairments in which they manifest themselves are of the sort that would standardly be explained at the personal level, rather than at the subpersonal level. In the terms of Fodor's dichotomy, psychiatric disorders seem to be disorders of central processing rather than peripheral modules …. Many of the symptoms of psychiatric disorders involve impairments of rationality—and consequently that the norms of rationality must be taken to play a vital role in the understanding of psychiatric disorders' (pp.460, 461).

Thus, there is an important sense in which rationality is a more encompassing construct than intelligence—it concerns both aspects of System 2. The reason is that rationality is an organismic-level concept. It concerns the actions of an entity in its environment that serve its goals. To be rational, an organism must have well calibrated beliefs (reflective level) and must act appropriately on those beliefs to achieve its goals (reflective level). The organism must, of course, have the algorithmic-level machinery that enables it to carry out the actions and to process the environment in a way that enables the correct beliefs to be fixed and the correct actions to be taken.

Thus, individual differences in rational thought and action can arise because of individual differences in intelligence or because of individual differences in thinking dispositions. To put it simply, the concept of rationality encompasses two things (thinking dispositions and algorithmic-level capacity) whereas the concept of intelligence—at least as it is commonly operationalized—is largely confined to algorithmic-level capacity.

Intelligence tests and critical thinking tests: Partitioning the algorithmic from the reflective mind

The difference between the reflective mind and the algorithmic mind is captured operationally in the distinction that psychologists make between tests of intelligence and tests of critical thinking. To a layperson, the tasks on tests of cognitive capacities (intelligence tests or other aptitude measures) might seem superficially similar to those on tests of critical thinking (in the educational literature, the term critical thinking is often used to cover tasks and mental operations that a cognitive scientist would term indicators of rational thought). An outsider to psychometrics or cognitive science might deem the classification of tasks into one category or the other somewhat arbitrary. In fact, it is far from arbitrary and actually reflects the distinction between the reflective mind and the algorithmic mind.

Psychometricians have long distinguished typical performance situations from optimal (sometimes termed maximal) performance situations (Ackerman and Heggestad, 1997; Ackerman and Kanfer, 2004). Typical performance situations are unconstrained in that no overt instructions to maximize performance are given, and the task interpretation is determined to some extent by the participant. In contrast, optimal performance situations are those where the task interpretation is determined externally (not left to the participant), the participant is instructed to maximize performance, and is told how to do so. All tests of intelligence or cognitive aptitude are optimal performance assessments, whereas measures of critical or rational thinking are often assessed under typical performance conditions. What this means is that tests of intelligence are constrained at the level of reflective processing (an attempt is made to specify the task demands so explicitly that variation in thinking dispositions is minimally influential). In contrast, tests of critical or rational thinking are not constrained at the level of reflective processing (or at least are much less constrained). Tasks of the latter but not the former type allow high-level personal goals (and epistemic goals) to become implicated in performance.

Consider the type of syllogistic reasoning item usually examined by cognitive psychologists studying belief bias effects (see Evans et al., 1983; Evans and Curtis-Holmes, 2005):

Premise 1: All living things need water

Premise 2: Roses need water

Therefore, Roses are living things

Approximately 70% of the university students who have been given this problem incorrectly think that the conclusion is valid (Markovits and Nantel, 1989; Sá et al., 1999;

Stanovich and West, 1998c). Clearly, the believability of the conclusion is interfering with the assessment of logical validity.

The important point for the present discussion is that it would not be surprising to see an item such as the 'rose' syllogism (that is, an item that pitted prior belief against logical validity) on a critical thinking test. Such tests do not constrain reflective-level thinking dispositions, and in fact attempt to probe and assess the nature of such cognitive tendencies to bias judgments in the direction of prior belief or to trump prior belief with new evidence (see, for example, certain exercises in the *Watson-Glaser Critical Thinking Appraisal*, Watson and Glaser, 1980).

In using items with such content, critical thinking tests create (even if the instructions attempt to disambiguate) ambiguity about what feature of the problem to rely upon—ambiguity that is resolved differently by individuals with different epistemic dispositions. The point is that on an intelligence test, there would be no epistemic ambiguity created in the first place. Such tests attempt to constrain reflective-level functioning in order to isolate processing abilities at the algorithmic level of analysis. It is the efficiency of computational abilities under optimal (not typical) conditions that is the focus of IQ tests. Variation in thinking dispositions would contaminate this algorithmic-level assessment.

I do not wish to argue that intelligence tests are entirely successful in this respect—that they entirely eliminate reflective-level factors; only that the constructors of the tests *attempt* to do so. Additionally, it is certainly the case that some higher-level strategic control is exercised on intelligence test items, but this tends to be a type of micro-level control rather than the activation of macro-strategies that are engaged by critical thinking tests. For example, on multiple-choice IQ-test items, the respondent is certainly engaging in a variety of control processes such as suppressing responses to identified distracter items. Nonetheless, if the test is properly designed, they are not engaging in the type of macro-level strategizing that is common on critical thinking tests—for example, deciding how to construe the task or how to allocate effort across differing construals.

Thus, you will not find an item like the 'rose' syllogism on an intelligence test (or any aptitude measure or cognitive capacity measure). For example, on a cognitive ability test, a syllogistic reasoning item will be stripped of content (all As are Bs, etc.) in order to remove any possible belief bias component. In complete contrast, in the reasoning and rational thinking literature, conflict between knowledge and validity is often deliberately *created* in order to study belief bias. Thus, cognitive ability tests eliminate the conflict between epistemic tendencies to preserve logical validity and the tendency to project prior knowledge. In contrast, critical thinking tasks deliberately leave reflective-level strategic decisions unconstrained, because it is precisely such epistemic regulation that they wish to assess. Of course this is why debates about the normative response on rational thinking measures have been prolonged in a way that has not characterized IQ tests (Cohen, 1981; Gigerenzer, 1996; Kahneman and Tversky, 1996; Manktelow, 2004; Over, 2002, 2004; Shafir and LeBoeuf, 2002; Stanovich, 1999; Stein, 1996). The more a measure taps the reflective-level psychology of rationality, the more it will implicate normative issues that are largely moot when measuring algorithmic-level efficiency.

The key functions of the reflective mind and the algorithmic mind that support human rationality

The reflective mind and the algorithmic mind both have a key function that serves to support human rationality. Both functions relate to an aspect of reasoning that has received considerable attention in parts of the dual-process literature—hypothetical thinking (Evans, 2003, 2006b, 2007, this volume; Evans and Over, 1996, 2004). One idea is that 'the analytic system is involved whenever hypothetical thought is required' (p.379, Evans, 2006b). Stated in the form of a conditional, we might say that: If hypothetical thought is required, then the analytic system is involved. Such a formulation preserves an important point I will make later—that not all analytic system thought involves hypothetical thinking.

Hypothetical thinking is the foundation of rationality because it is tightly connected to the notion of TASS override (see Stanovich, 2004). The analytic system must be able to take early response tendencies triggered by TASS offline and be able to substitute better responses. But where do these better responses come from? One answer is that they come from a process of cognitive simulation (e.g. Buckner and Carroll, 2007; Byrne, 2005; Kahneman and Tversky, 1982b; Nichols and Stich, 2003; Oatley, 1999). Responses that have survived a selective process during simulation are often a better choice than the TASS-triggered response. So the key mechanism of the *reflective* mind that supports human rationality is the mechanism that sends out a call to begin cognitive simulation or hypothetical reasoning more generally. It is conjectured that individual differences in the operation of this mechanism contribute to the differences in rational thinking examined in some of the studies cited above (e.g. Stanovich and West, 1998c).

Correspondingly, there is a key operation of the algorithmic mind that supports hypothetical thinking and that is characterized by large individual differences. Simply put, cognitive simulation and hypothetical reasoning is dependent upon the operation of cognitive decoupling carried out by the algorithmic mind. Cognitive decoupling has been discussed in related and somewhat differing ways by a large number of different investigators coming from a variety of different perspectives, not limited to: developmental psychology, evolutionary psychology, artificial intelligence, and philosophy of mind (Cosmides and Tooby, 2000; Dienes and Perner, 1999; Jackendoff, 1996; Nichols and Stich, 2003; Perner, 1991; Sperber, 2000). I shall emphasize the origins of the concept in developmental psychology because of a useful theoretical link to important models of the origins of System 2 (see Mithen, 1996).

In a famous article in the early theory of mind literature, Leslie (1987) provided a model of pretence that made use of the concept of cognitive decoupling. Leslie's (1987) model can best be understood by adopting a terminology later used by Perner (1991). In the latter's view, a primary representation is one that is used to directly map the world and/or is also rather directly connected to a response. Leslie (1987) modeled pretense by positing a so-called secondary representation (to use Perner's [1991] terms) that was a copy of the primary representation but that was decoupled from the world so that it could be manipulated—that is, be a mechanism for simulation. Nichols and Stich (2003) model this cognitive decoupling as a separate 'possible world box' (PWB) in which the simulations are carried out without contaminating the relationship between the world and primary representation.

For Leslie (1987), the decoupled secondary representation is necessary in order to avoid so-called representational abuse—the possibility of confusing our simulations with our primary representations of the world as it actually is. The cognitive operation of decoupling, or what Nichols and Stich (2003) term cognitive quarantine, prevents our representations of the real world from becoming confused with representations of imaginary situations. For example, when considering an alternative goal state different from the current goal state, one needs to be able to represent both. To engage in these exercises of hypotheticality and high-level cognitive control, one has to explicitly represent a psychological attitude toward the state of affairs as well as the state of affairs itself. Thus, decoupled representations of actions about to be taken become representations of potential actions, but the latter must not infect the former while the mental simulation is being carried out.

Decoupling operations must be continually in force during any ongoing simulations, and I have conjectured (Stanovich, 2001, 2004) that the raw ability to sustain such mental simulations while keeping the relevant representations decoupled is likely the key aspect of the brain's computational power that is being assessed by measures of fluid intelligence (on fluid intelligence, see Carroll, 1993; Horn and Noll, 1997; Kane and Engle, 2002).

Decoupling—outside of certain domains such as behavioral prediction (so-called 'theory of mind' where evolution has built content-specific machinery)—is a cognitively demanding operation. Any mindware that can aid this computationally expensive process is thus immensely useful, and language appears to be one such mental tool. Language provides the discrete representational medium that greatly enables hypotheticality to flourish as a culturally acquired mode of thought. For example, hypothetical thought involves representing assumptions, and linguistic forms such as conditionals provide a medium for such representations (Carruthers, 2002; Evans, 2007; Evans and Over, 2004).

Decoupling skills vary in their recursiveness and complexity. The skills discussed thus far are those that are necessary for creating what Perner (1991) calls secondary representations—the decoupled representations that are the multiple models of the world that enable hypothetical thought. At a certain level of development, decoupling becomes used for so-called meta-representation—thinking about thinking itself (there are many subtleties surrounding the concept of metarepresentation; see Dennett, 1984; Perner, 1991; Sperber 2000; Whiten, 2001). Decoupling processes enable one to distance oneself from representations of the world so that they can be reflected upon and potentially improved. The use of metarepresentational abilities in such a program of cognitive reform would be an example of what has been termed the quest for broad rationality—the cognitive critique of the beliefs and desires that are input into the implicit calculations that result in instrumental (Humean) rationality (see Stanovich, 2004).

I propose that cognitive decoupling is the key function of the algorithmic mind that supports human rationality and that it is the operation that accounts for several other features of what we have been calling System 2—in particular its seriality and most importantly its computational expense. In short, we are beginning to understand the key computational function of the algorithmic mind—which is to maintain decoupling among representations while carrying out mental simulation. This is becoming clear from converging work on executive function (Baddeley et al., 2001; Duncan, et al., 2000; Hasher et al., 1999; Kane, 2003; Kane and Engle, 2002; Salthouse et al., 2003) and working memory (Conway et al., 2003; Engle, 2002; Geary, 2005; Kane et al., 2001; Kane and Engle, 2003; Kane et al., 2005).

First, there is a startling degree of overlap in individual differences on working memory tasks and individual differences in measures of fluid intelligence. Secondly, it is becoming clear that working memory tasks are only incidentally about memory. Or, as Engle (2002) puts it,

> 'WM capacity is just as important in retention of a single representation, such as the representation of a goal or of the status of a changing variable, as it is in determining how many representations can be maintained. WM capacity is not directly about memory—it is about using attention to maintain or suppress information. WM capacity is about memory only indirectly. Greater WM capacity does mean that more items can be maintained as active, but this is a result of greater ability to control attention, not a larger memory store' (p.20).

Hasher et al.(2007) concur with this view when they conclude that 'our evidence raises the possibility that what most working memory span tasks measure is inhibitory control, not something like the size of operating capacity' (p.231).

Lepine et al. (2005) report an experiment showing that working memory tasks with simple processing components are actually better predictors of high-level cognitive performance than are working memory tasks with complex processing requirements—as long as the former are rapidly paced to lock up attention. Their results are consistent with Engle's (2002) review of evidence indicating that working memory tasks really tap the preservation of internal representations in the presence of distraction or, as I have termed it—the ability to decouple a secondary representation (or metarepresentation) from a primary representation and manipulate the former. For example, he describes an experiment using the so-called antisaccade task. Subjects must look at the middle of a computer screen and respond to a target stimulus that will appear on the left or right of the screen. Before the target appears, a cue is flashed on the opposite side of the screen. Subjects must resist the attention-capturing cue and respond to the target on the opposite side when it appears. Subjects scoring low on working memory tasks were more likely to make an eye movement (saccade) in the direction of the distracting cue than were subjects who scored high on working memory task.

That the antisaccade task has very little to do with memory is an indication of why investigators have reconceptualized the individual difference variables that working memory tasks are tapping. Individual differences on such tasks are now described with a variety of different terms (attentional control, resistance to distraction, executive control), but the critical operation needed to succeed in them—and the reason they are the prime indicator of fluid intelligence—is that they reflect the ability to sustain decoupled representations. Such decoupling is an important aspect of behavioral control that is related to rationality (see De Neys, 2006a, 2006b).

So-called 'executive functioning' measures tap the algorithmic mind and not the reflective mind

One interesting implication that follows from the distinction between the algorithmic mind and reflective mind is that the measures of so-called executive functioning in the neuropsychological literature actually measure nothing of the sort. The term 'executive' implies that these tasks assess the highest level of cognitive functioning—the reflective level. However, a consideration of the tasks most commonly used in the neuropsychological literature to assess executive functioning (see Pennington and Ozonoff, 1996; Salthouse et al., 2003) reveals that almost without exception they are optimal performance tasks and not typical performance tasks and that most of them rather severely constrain intentional-level functioning. Thus, because intentional-level functioning is constrained, such tasks are largely assessing individual differences in algorithmic-level functioning. This is the reason why several studies have shown very strong correlations between executive functioning and fluid intelligence (Conway et al., 2003; Kane et al., 2005; Salthouse et al., 2003; Unsworth and Engle, 2005).

Consider some of the classic tasks in the neuropsychological literature on executive function (see Pennington and Ozonoff, 1996; Salthouse et al., 2003). In the critical part of the Trail Making Test the subject must, in the shortest time possible, connect with a line a series of numbered and lettered circles going from 1 to A to 2 to B to 3 to C, etc. The rule is specified in advance and there is no ambiguity about what constitutes optimal performance. There is no higher-level task interpretation required of the subject. Cognitive decoupling is required though, in order to keep the right sequence in mind and not revert to number sequencing alone or letter sequencing alone. Thus, the task does require algorithmic-level decoupling in order to suppress TASS from disrupting performance by defaulting to an overlearned rule. But the task does not require reflective control in the sense that I have defined it here (or it does in only the most basic sense by requiring a decision to comply with the tester or experimenter).

The situation is similar regarding another test of executive functioning from the neuropsychological literature, the Stroop Test. The subject is explicitly told to name the color and not read the word, and optimal performance is clearly defined as going as fast as possible. Algorithmic-level decoupling is needed in order to suppress the automatic response from TASS to read the word. But higher-level reflective control never enters the picture. The response requirements of the task are very basic and the task set is dictated externally. It is a test of suppression via algorithmic-level decoupling pure and simple. Fluency tasks are also commonly used to measure executive functioning (Salthouse et al., 2003). Here, the subject simply articulates as many words as they can from a specified category (words beginning with the letter F, names of red things, etc.). Again, in such a task there is no reflective choice about what rule to use. The task requirements are entirely specified in advance and the assessment concerns merely the efficiency of execution.

A widely used measure of executive functioning, the Wisconsin Card Sorting Test (Heaton et al., 1993), does begin to tap more reflective processes, although variance in suppression via decoupling is still probably the dominant individual difference component that it taps. In the WCST the subject sees a set of target cards containing shapes varying in color, form, and number. The instructions are to sort new cards in a deck correctly by grouping them with the correct target card. The subject must discover the dimension (color, form, or number) that should be the basis of the sort, and at predetermined points the correct dimension of sort is changed on the subject without warning. Although the basic task structure is set by the examiner, there may well be some reflective involvement in the rule discovery stages of the task. Nevertheless, once the rule is switched, suppression of the tendency to sort by the previous rule is probably the dominant influence on performance. This suppression is carried out by algorithmic-level decoupling abilities and is probably why the task is correlated with fluid intelligence (Salthouse et al., 2003).

The tasks I have discussed so far come from the neuropsychological literature. However, more precise experimental tasks have been used in the literature of cognitive psychology to measure exactly the same construct as the neuropsychological executive function measures. These more precise tasks—stop signal paradigms, working memory paradigms, time sharing paradigms, inhibition paradigms of various types (see Salthouse et al., 2003)—are all subject to exactly the same arguments just made regarding the neuropsychological measures. The more precise experimental measures are optimal performance tasks (not typical performance tasks) and they severely constrain reflective-level functioning. All measure algorithmic-level decoupling power, which is why they display a considerable degree of overlap with fluid intelligence (Kane and Engle, 2002; Salthouse et al., 2003). Individual differences in the reflective mind are only tangentially implicated. This is because tapping reflective processes requires measures of typical performance so that individual differences in epistemic regulation and cognitive allocation (e.g. need for cognition) become implicated in performance beyond simply the computational power to sustain decoupling operations. This point about the laboratory measures has been made before by Salthouse et al. (2003): 'The role of executive functioning may also be rather limited in many laboratory tasks because much of the organization or structure of the tasks is provided by the experimenter and does not need to be discovered or created by the research participant' (p.569).

In short, my argument is that executive processes are misnamed in the psychological literature. Executive functioning measures are nothing of the kind—at least as most people would understand the word 'executive'. These tasks might instead be better termed measures of *supervisory* processes. They assess the ability to carry out the rules instantiated not by internal regulation (*true* executive control) but by an external authority that explicitly sets the rules and tells the subject what constitutes maximal performance. Subjects do not set the agenda in these tasks (as is the case in many tasks in the rational thinking and critical thinking literatures) but instead attempts to optimize criteria explicitly given to them. The processes assessed by such tasks do involve algorithmic-level decoupling (which is why they are so highly related to fluid intelligence), but they are supervisory in nature—decoupling is used to screen out distracting stimuli (i.e. suppress via decoupling irrelevant inputs from the autonomous mind) and make sure the externally-provided rule remains the goal state.

In contrast, processes of the reflective mind operate to set the goal agenda or they operate in the service of epistemic regulation (i.e. to direct the sequence of information pickup). Such processes that set and regulate the goal and epistemic agendas are little engaged by so-called executive function tasks. The term 'executive' thus can lead to theoretical confusion in the literature. More importantly, it contributes to the tendency to overlook the importance of measuring variation the reflective mind. The term 'executive' mistakenly implies that everything 'higher up' has been taken care of, or that there is no level higher than what these executive functioning tasks measure.

Serial associative cognition with a focal bias

The current tripartite view of the mind has begun to look somewhat like that displayed in Figure 3.2. Previous dual-process theories have emphasized the processing sequence where the reflective mind sends out a call to the algorithmic mind to override the TASS response by taking it offline. An alternative response that is the result of cognitive simulation is substituted for the TASS response that would have been emitted. The override function has loomed large in dual-process theory but less so the simulation process that computes the alternative response that makes the override worthwhile. Figure 3.2 explicitly represents the simulation function as well as the fact that the call to initiate simulation originates in the reflective mind. The decoupling operation itself is carried out by the algorithmic mind. Recall that two different types of individual differences are associated with the initiation call and the decoupling operator—specifically, rational thinking dispositions with the former and fluid intelligence with the latter.

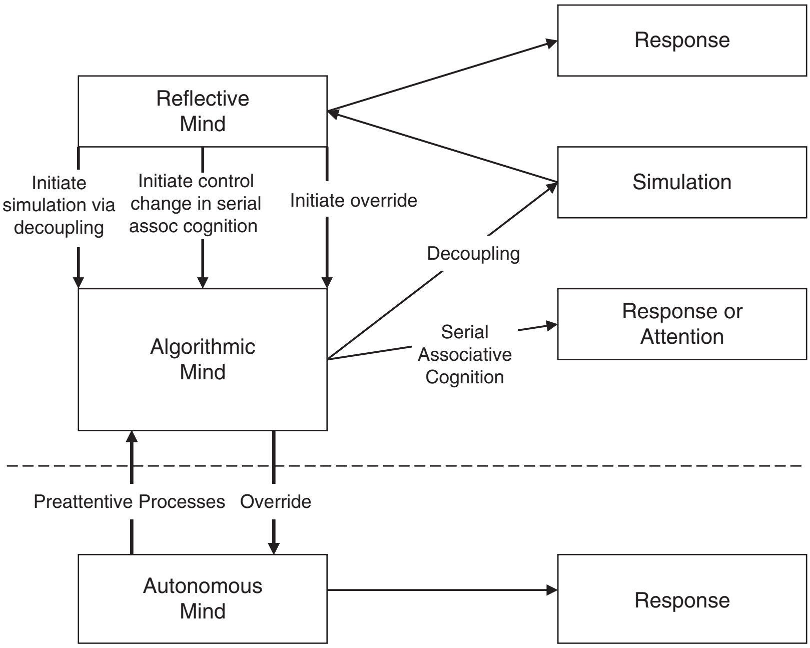


Fig. 3.2 A more complete model of the tripartite structure.

The model in Figure 3.2 defines a third critical function for the algorithmic mind in addition to TASS override and enabling simulation. The third is a function that in the Figure is termed serial associative cognition. This function relates to my point mentioned previously, that: All hypothetical thinking involves the analytic system (Evans and Over, 2004), but not all analytic system thought involves hypothetical thinking. Serial associative cognition represents this latter category. It can be understood by considering a discussion of the selection task in a recent theoretical paper on dual-processes by Evans (2006b; see also Evans and Over, 2004). Here, and in Evans and Over (2004), it is pointed out that the previous emphasis on the matching bias evident in the task (Evans, 1972, 1998, 2002; Evans and Lynch, 1973) might have led some investigators to infer that the analytic system is not actively engaged in the task. In fact, matching bias might be viewed as just one of several such suggestions in the literature that much thinking during the task is non-analytic (see Margolis, 1987; Stanovich and West, 1998a; Tweney and Yachanin, 1985). In contrast, however, Evans (2006b) presents evidence indicating that there may be analytic system involvement during the task—even on the part of the majority who do not give the normatively correct response but instead give the PQ response.

First, in discussing the card inspection paradigm (Evans, 1996) that he pioneered (see also Lucas and Ball, 2005; Roberts and Newton, 2001), Evans (2006b) notes that although subjects look disproportionately at the cards they will choose (the finding leading to the inference that heuristic processes were determining the responses), the lengthy amount of time they spend on those cards suggests that analytic thought is occurring (if only to generate justification for the heuristically-triggered choices). Secondly, in verbal protocol studies, subjects can justify their responses (indeed, can justify *any* set of responses they are told are correct, see Evans and Wason, 1976) with analytic arguments—arguments that sometimes refer to the hidden side of cards chosen.

I think it is correct to argue that analytic cognition is occurring in this task, but I also want to argue that it is not full blown cognitive simulation of alternative world models. It is thinking of a shallower type (see Frankish, 2004). In Figure 3.2 I have termed it serial associative cognition—cognition that is not rapid and parallel such as TASS processes, but is nonetheless rather inflexibly locked into an associative mode that takes as its starting point a model of the world that is *given* to the subject. In the inspection paradigm, subjects are justifying heuristically-chosen responses (P and Q for the standard form of the problem), and the heuristically-chosen responses are driven by the model given to the subject by the rule.

Likewise, Evans and Over (2004) note that in the studies of verbal protocols subjects making an incorrect choice referred to the hidden sides of the cards they are going to pick, but referred only to verification when they did so. Thus, the evidence suggests that subjects accept the rule as given, assume it is true, and simply describe how they would go about verifying it. The fact that they refer to hidden sides does not mean that they have constructed any alternative model of the situation beyond what was given to them by the experimenter and their own assumption that the rule is true. They then reason from this single focal model—systematically generating associations from this focal model but never constructing another model of the situation.

This is what I would term serial associative cognition with a focal bias. It is how I would begin to operationalize the satisficing bias in the analytic system posited by Evans (2006b, 2007; Evans et al., 2003).

One way in which to contextualize the idea of focal bias is as the second stage in a framework for thinking about human information processing that is over 30 years old—the idea of humans as cognitive misers (Dawes, 1976; Taylor, 1981; Tversky and Kahneman, 1974). Krueger and Funder (2004) characterize the cognitive miser assumption as one that emphasizes 'limited mental resources, reliance on irrelevant cues, and the difficulties of effortful correction' (pp.316–7). More humorously, Hull (2001) has said that 'the rule that human beings seem to follow is to engage the brain only when all else fails—and usually not even then' (p.37).

There are in fact several aspects of cognitive miserliness. Dual-process theory has heretofore highlighted only Rule 1 of the Cognitive Miser: default to TASS processing whenever possible. But defaulting to TASS processing is not always possible—particularly in novel situations where there are no stimuli available to domain-specific evolutionary modules, nor perhaps any information with which to run overlearned and well-compiled procedures that TASS has acquired through practice. Analytic processing procedures will be necessary, but a cognitive miser default is operating even there. Rule 2 of the Cognitive Miser is that, when analytic processing is necessary: default to serial associative cognition with a focal bias (*not* fully decoupled cognitive simulation). [Rule 3 might be deemed the tendency to start cognitive simulation but not complete it—that is, override failure.]

Evans (2006b) draws attention to Rule 2 in the model of humans as cognitive misers by emphasizing a satisficing principle in his conception of the analytic system. The notion of focal bias is a way of conceiving of just what satisficing by the analytic system is in terms of actual information processing mechanics. The proposal is, simply, that it amounts to a focal bias with an additional tendency *not* to interrupt serial associative cognition with a decoupling call from the reflective mind.

The notion of a focal bias conjoins several closely related ideas in the literature—Evans et al.'s (2003) singularity principle, Johnson-Laird's (1999, 2005) principle of truth, focusing (Legrenzi et al., 1993), the effect/effort issues discussed by Sperber et al. (1995), and finally, the focalism (Wilson et al., 2000) and belief acceptance (Gilbert, 1991) issues that have been prominent in the social psychological literature. My notion of focal bias conjoins many of these ideas under the overarching theme that they all have in common—that humans will find any way they can to ease the cognitive load and process less information. Focal bias combines all of these tendencies into the basic idea that the information processor is strongly disposed to deal only with the most easily constructed cognitive model.

So the focal model that will dominate processing—the only model that serial associative cognition deals with—is the most easily constructed model. The focal model tends to represent: only one state of affairs (the Evans et al., 2003, singularity idea), it accepts what is directly presented and models what is presented as true (e.g. Gilbert, 1991; Johnson-Laird, 1999), it is a model that minimizes effort (Sperber et al., 1995), it ignores moderating factors (as the social psychological literature has demonstrated, e.g. Wilson, et al., 2000)—probably because taking account of those factors would necessitate modeling several alternative worlds and this is just what a focal processing allows us to avoid. And finally, given the voluminous literature in cognitive science on belief bias and the informal reasoning literature on my side bias, the easiest models to represent clearly appear to be those closest to what a person already believes in and has modeled previously (e.g. Evans and Feeney, 2004; Stanovich and West, 2007).

Thus, serial associative cognition is defined by its reliance on a single focal model that triggers all subsequent thought. So framing effects, for instance, are a clear example of serial associative cognition with a focal bias. As Kahneman (2003) notes, 'the basic principle of framing is the passive acceptance of the formulation given' (p.703). The frame presented to the subject is taken as focal, and all subsequent thought derives from it rather than from alternative framings because the latter would necessitate more computationally expensive simulation operations.

In short, serial associative cognition is serial and analytic (as opposed to holistic) in style, but it relies on a single focal model that triggers all subsequent thought. Such a view is consistent with the aforementioned discussion of thinking during the selection task and the conclusion that analytic cognition does indeed take place even for the incorrect responders (see Evans, 2006b; Evans and Over, 2004). Incorrect responders are engaging in serial associative cognition with a focal bias, but reflective processes are not prone to send additional decoupling calls in order to explore alternative models to the focal one. A final factor that might differentiate serial associative cognition from fully decoupled simulation is the tendency for the focal model in the former to become 'unclamped'—that is, to be replaced by another model suggested by the serial stream of consciousness.

In the tripartite model proposed here, the decoupling operation is uniquely a function of the algorithmic mind—it is not a function of TASS (outside of the theory of mind module). It is also the main source of variance in computational assessments of the algorithmic mind (such as tests of fluid intelligence). But again I would stress that what is assessed on such measures is the ability to *sustain* cognitive decoupling when the necessity for decoupling is clearly communicated to the subject. Such measures do not in fact assess the natural *tendency* to simulate alternative models—they do not assess the tendency of the reflective mind to send out an instruction to decouple from the focal model.

The preceding discussion might be taken to define three different functions of cognitive decoupling. In the override case, decoupling involves taking offline the connection between a primary representation and response programming in TASS. In the second case, the case of comprehensive simulation, it involves segregating from representational abuse multiple models undergoing simultaneous evaluation and transformation. Of course these two are related—TASS responses are often decoupled pending a comprehensive simulation that determines whether there is a better response.

A third type of decoupling involves interrupting serial associative cognition—that is, decoupling from the next step in an associative sequence that would otherwise direct thought. This third type of decoupling might shunt the processor to comprehensive simulation or may not in fact replace the focal model but simply start a new associative chain from a different starting point *within* the focal model.

Paralleling the three types of decoupling are three initiate signals from the reflective mind (see Figure 3.2): initiating override operations; initiating cognitive simulation; and initiating an interrupt of serial associative cognition.[2]

Dual-process theory and knowledge structures

One aspect of dual process theory that has been relatively neglected is that the simulation process is not simply procedural but instead utilizes content—that is, it uses declarative knowledge and strategic rules (linguistically-coded strategies) to transform a decoupled representation. In the previous dual-process literature, override and simulation have been treated as somewhat disembodied processes. The knowledge bases and strategies that are brought to bear on the secondary representations during the simulation process have been given little attention.

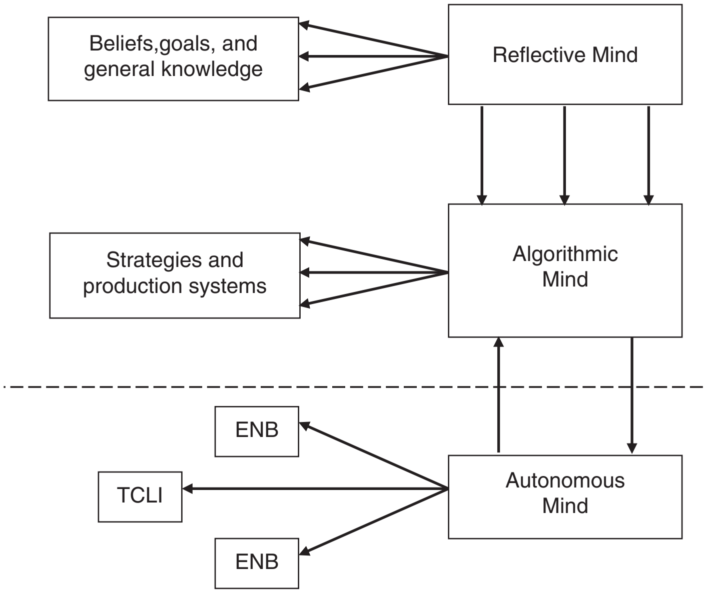
In fact, each of the levels in the tripartite model described in this chapter has to access knowledge to carry out its operations (see Figure 3.3). The reflective mind not only accesses general knowledge structures but, importantly, accesses the person's opinions, beliefs, and reflectively acquired goal structure (considered preferences, see Gauthier, 1986). The algorithmic mind accesses micro-strategies for cognitive operations and production system rules for sequencing behaviors and thoughts. Finally, the autonomous mind accesses not only evolutionarily-compiled encapsulated knowledge bases, but also retrieves information that has become tightly compiled due to overlearning and practice.

It is important to note that what is displayed in Figure 3.3 are the knowledge bases that are *unique* to each mind. Algorithmic- and reflective-level processes also receive inputs from the computations of the autonomous mind. As Evans (this volume) notes, TASS processes that supply information to the analytic system are sometimes termed preattentive processes. His chapter contains a particularly good discussion of the importance of these preattentive processes in fixing the content of analytic thought.

The rules, procedures, and strategies that can be retrieved by the analytic system (the algorithmic and reflective minds) and used to transform decoupled representations have been referred to as mindware, a term coined by David Perkins in a 1995 book (Clark, 2001, uses it a slightly different way from Perkins' original coinage). The mindware available for the analytic system to substitute during TASS override is in part the product of past learning experiences. Indeed, if one is going to trump a TASS-primed response with conflicting information or a learned rule, one must have previously learned the information or the rule. If, in fact, the relevant mindware is not available because it has not been learned, then we have a case of missing mindware rather than a TASS-override failure. This distinction in fact represents the beginning of a taxonomy of the causes of cognitive failure related to rational behavior that I am currently using to organize the heuristics and biases literature and to classify various practical problems of rational thinking—for example, to understand the thinking problems of pathological gamblers (Toplak et al., 2007).

---

[2] These three functions of decoupling are interestingly parallel to three executive process functions (see Miyake et al., 2000) that have been discussed in the literature: inhibition, updating, and set shifting.



ENB = Encapsulated Knowledge Base
TCLI = Tightly Compiled Learned Information

Fig. 3.3 Knowledge structures in the tripartite model.

A taxonomy applied to the heuristics and biases literature

My taxonomy recognizes several different categories of cognitive failure termed: TASS override failure; mindware gaps; contaminated mindware; defaulting to the autonomous mind, and defaulting to serial associative cognition with a focal bias.[3] The first is the well-known category—the one we are all familiar with from the dual-process literature: situations where the TASS-primed responses must be

---

[3] I have discussed a fifth category elsewhere (Stanovich, in press b; Toplak et al., 2007), but because it relates less to classifying the heuristics and biases literature, I omit it here. The fifth category derives from the possibility of not too much TASS output (as in override failure) but too little. Cognitive neuroscientists have uncovered cases of mental pathology that are characterized by inadequate behavioral regulation from the emotion modules in TASS—for example, Damasio's (1994, 1996; Bechara et al., 1994; Eslinger and Damasio, 1985) well-known studies of patients with damage in the ventromedial prefrontal cortex. These individuals have severe difficulties in real-life decision making but do not display the impairments in sustained attention and executive control that are characteristic of individuals with damage in dorsolateral frontal regions (e.g. Bechara, 2005; Duncan et al., 1996; Kimberg et al,, 1998; McCarthy and Warrington, 1990; Pennington and Ozonoff, 1996). Instead, they are thought to lack the emotions that constrain the combinatorial explosion of possible actions to a manageable number based on somatic markers stored from similar situations in the past.

overridden by the analytic system if the optimal response is to be made and the analytic system fails to override. As just mentioned though, this category is interestingly related to the second. Note that there are two reasons for what previously has been termed a failure of TASS override, but most discussions in the dual-process literature simply tacitly default to one of them. Most previous discussions of TASS-override have simply assumed that mindware was available to be employed in an override function by the analytic system. If, in fact, the mindware is not available because it has not been learned or at least not learned to the requisite level to sustain override, then I am suggesting in this taxonomy that we call this not override failure but instead a mindware gap.

Note one interesting implication of the relation between TASS override and mindware gaps—the fewer gaps one has, the more 'at risk' one will be for a case of override failure. Someone with considerable mindware installed will be at greater risk of failing to use it in a propitious way. Of course, the two categories trade off in a continuous manner with a fuzzy boundary between them. A well-learned rule not appropriately applied is a TASS override failure. As the rule is less and less well instantiated, at some point it is so poorly compiled that it is not a candidate to override the TASS response and thus the processing error becomes a mindware gap. The study of pathological gambling behavior, for instance, has focused on a class of missing mindware of particular relevance to that condition: knowledge and procedures for dealing with probability and probabilistic events (Keren, 1994; Toplak et al., 2007; Wagenaar, 1988). Many studies now administer to such subjects measures of knowledge of regression to the mean, outcome bias, covariation detection, the gambler's fallacy, probability matching, baserate neglect, Bayesian probabilistic updating, and covariation detection.

Although mindware gaps may lead to sub-optimal reasoning, the next category in the taxonomy is designed to draw attention to the fact that not all mindware is helpful—either to goal attainment or to epistemic accuracy. In fact, some acquired mindware can be the direct cause of irrational actions that thwart our goals. Such effects thus define another category in the taxonomy of cognitive failures: contaminated mindware. Although the idea of contaminated mindware is controversial (see Aunger, 2000) many theorists speculating on the properties of cultural replication would admit such a possibility (Blackmore, 1999, 2005; Dennett, 1991, 2006; Distin, 2005; Hull, 2000; Mesoudi et al., 2006).

Two further categories are defined by two effort-minimizing strategies that reflect cognitive miserliness. One is the tendency to engage in serial associative cognition with a focal bias (see Evans, 2006b, on satisficing during analytic processing). This represents a tendency to over-economize during analytic processing—specifically, to fail to engage in the full-blown simulation of alternative worlds or to engage in fully disjunctive reasoning (Shafir, 1994; Toplak and Stanovich, 2002). Another effort minimizing strategy is the tendency not to engage in System 2 reasoning at all (not even serial associative cognition)—specifically, to default to the processing options offered by the autonomous mind.

Figure 3.4 displays the different classes of cognitive dysfunction. Many useful distinctions are captured in the Figure. For example, it helps to illustrate the importance of distinguishing between TASS override failure and situations where System 2 is not engaged at all. In my use of the term, for something to be considered a TASS override

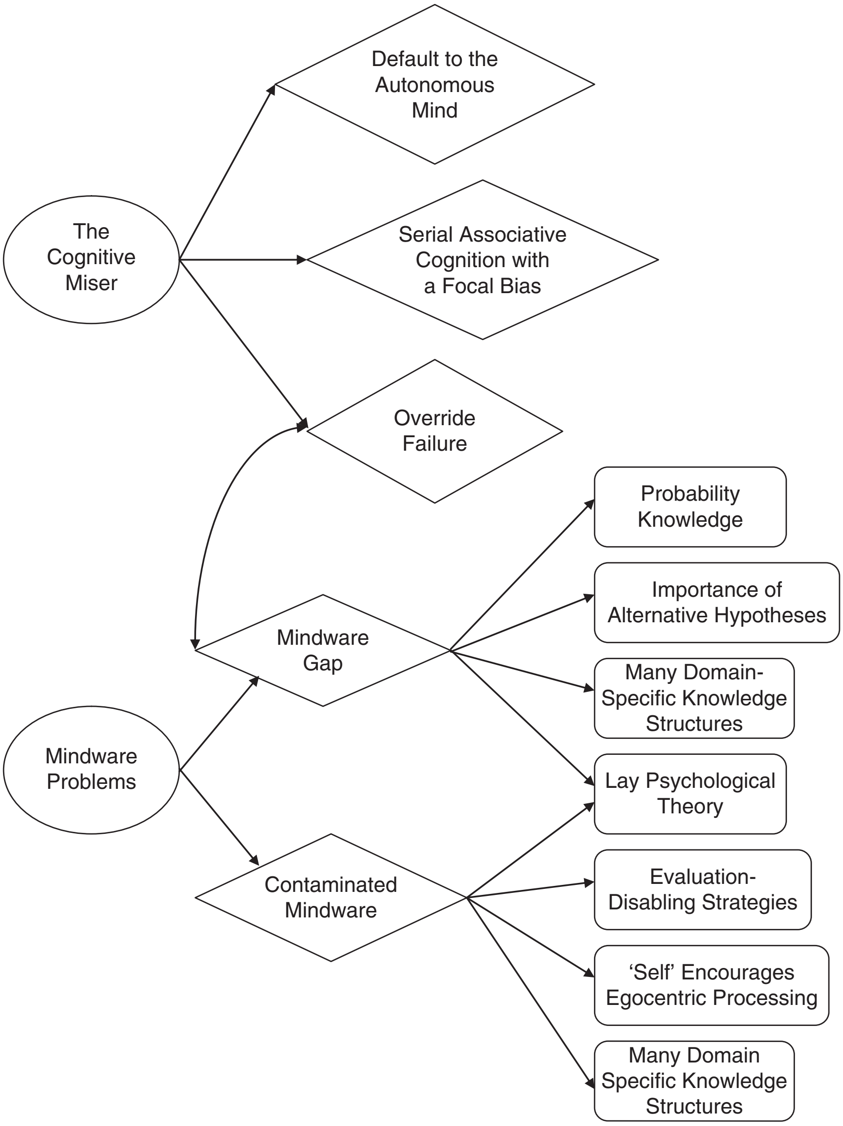


Fig. 3.4 A basic taxonomy of thinking errors.

failure, the analytic system must lose in a conflict of discrepant outputs. If the analytic system is not engaged at all, then we have a case of defaulting to the autonomous mind. In fact, the early heuristics and biases researchers were clearer on this point than many later dual-process theorists. The distinction between impressions and judgments in the early heuristics and biases work (see Kahneman, 2003; Kahneman and Frederick, 2002, 2005, for a discussion) made it clearer that non-normative responses often resulted not from a TASS/System 2 struggle but from intuitive impressions that are left uncorrected by System 2 rules and strategies. In fact, in many cases that have been called TASS override failure in the literature, the subject probably does not even consider overriding the TASS-based response (even when the mindware to do so is readily available and well learned). The subject does not recognize the need for override, or chooses not to sustain the necessary decoupling and simulation with alternative mindware that would make override possible.

The category of true override failure in my taxonomy would encompass what folk theory would call problems of willpower or the problem of multiple minds (see Ainslie, 2001, for a nuanced discussion of this folk concept in light of modern cognitive science). But there is more than just willpower issues in this category. Heuristics and biases tasks can also trigger the problem of multiple minds. Sloman (1996) points out that at least for some subjects, the Linda conjunction problem (see Tversky and Kahneman, 1983) is the quintessence of dual-process conflict. He quotes Stephen Gould's introspection that 'I know 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 be a bank teller; read the description"' (Gould, 1991, p.469). For sophisticated subjects such as Gould, resolving the Linda problem clearly involves a TASS/analytic system conflict, and in his case a conjunction error on the task would represent a true TASS override failure. However, for the majority of subjects, there is no conscious introspection going on—System 2 is either not engaged or engaged so little that there is no awareness of a cognitive struggle. Instead, TASS-based heuristics such as representativeness or conversational pragmatics trigger the response (the detailed controversies about the Linda task are beyond the scope of the present chapter; for that large literature, see Adler, 1984, 1991; Girotto, 2004; Lee, 2006; Mellers et al., 2001; Politzer and Macchi, 2000). Analytic processing has not lost a struggle when it has not been called into the battle.

Defaulting to the autonomous mind is a more miserly type of thinking error than is override failure. In the former, sustained decoupling is not even attempted, whereas in the latter decoupling is initiated but is not sustained until completion. Intermediate between these two is System 2 processing taking place without sustained decoupling: serial associative cognition with a focal bias. In Figure 3.4 it is displayed between the other two. Figure 3.4 thus displays three types of cognitive miserliness and, below them, two types of mindware problem. The first mindware category in Figure 3.4 is the category of mindware gaps. The double-headed arrow indicates its aforementioned relation to TASS override. Some representative areas where important mindware gaps occur are illustrated. I have not represented an exhaustive set of knowledge partitionings—to the contrary, I have represented a minimal sampling of a potentially large set of coherent knowledge bases in the domains of probabilistic reasoning, causal reasoning, logic, and scientific thinking, the absence of which could result in irrational thought or behavior. I have represented mindware categories that have been implicated in research in the heuristics and biases tradition: missing knowledge about probability and probabilistic reasoning strategies; and ignoring alternative hypotheses when evaluating hypotheses. The latter would encompass the phenomenon of evaluating hypotheses in a way that implies that one is ignoring the denominator of the likelihood ratio in Bayes' rule—the probability of D given ~H [P(D/~H)]. These are just a few of many mindware gaps that have been suggested in the literature on behavioral decision making. There are many others, and the box labeled 'Many Domain-Specific Knowledge Structures' indicates this.

Finally, at the bottom of the Figure is the category of contaminated mindware. Again, the curved rectangles do not represent an exhaustive partitioning (the mindware-related categories are too diverse for that), but instead represent some of the mechanisms that have received some discussion in the literature. One is a subcategory of contaminated mindware that is much discussed in the memetics literature—memeplexes that contain evaluation-disabling memes (Blackmore, 1999; Dennett, 1991, 2006; Lynch, 1996; Stanovich, 2004). Some of the evaluation-disabling memes that help keep some memeplexes lodged in their hosts are: memes that promise punishment if the memeplex is questioned; those that promise rewards for unquestioning faith in the memeplex; or those that thwart evaluation attempts by rendering the memeplex unfalsifiable.

Another subcategory of contaminated mindware has been suggested by memetic theorists such as Dennett (1991, 1995) and Blackmore (1999) who consider the self to be a memetic construct. Among its many properties is the fact that the self serves to encourage egocentric thinking. Thus, the self is a mechanism that fosters one characteristic of focal bias: that we tend to build models of the world from a single my side perspective. Nevertheless, if should not be forgotten that the egocentrism of memeplex self must be a very adaptive cognitive style—both evolutionarily adaptive and adaptive in the sense of our personal (that is, vehicle) goals. However, for many of the same reasons that TASS heuristics often are non-optimal in a technological environment different from the environment of evolutionary adaptation, the decontextualizing demands of modernity increasingly require such characteristics as: fairness, rule-following despite context, even-handedness, nepotism prohibitions, unbiasedness, universalism, inclusiveness, contractually mandated equal treatment, and discouragement of familial, racial, and religious discrimination. These requirements are difficult ones probably for the reason that they override processing defaults related to the self.

Finally, the last subcategory of contaminated mindware pictured in Figure 3.4 (labeled Many Domain-Specific Knowledge Structures) is meant to represent what is actually a whole set of categories: mindware representing specific categories of information or maladaptive memeplexes. Like the missing mindware category, there may be a large number of misinformation-filled memeplexes that would support irrational thought and behavior. For example, the gambler's fallacy and many of the other misunderstandings of probability that have been studied in the heuristics and biases literature would fit here. Of course, this example highlights the fact that the line between missing mindware and contaminated mindware might get fuzzy in some cases and the domain of probabilistic thinking is probably one such case.

Problems with people's lay psychological theories are represented as both contaminated mindware *and* a mindware gap in Figure 3.4. Mindware gaps are the many things about our own minds that we do not know; for example, how quickly we will adapt to both fortunate and unfortunate events (Gilbert, 2006). Other things we think we know about our own minds are wrong. These misconceptions represent contaminated mindware. An example would be the folk belief that we accurately know our own minds. This contaminated mindware accounts for the incorrect belief that we always know the causes of our own actions (Nisbett and Wilson, 1977) and the tendency to think that although others display thinking biases, we ourselves have special immunity from the very same biases (Pronin, 2006).

Finally, Table 3.1. illustrates how the various cognitive characteristics and processing styles that exemplify the category cash out in terms of well-known effects and tasks in the thinking and reasoning literature. This again is not an exhaustive list, and I have detailed some categories more than others to reflect my interests and the biases of the field. The grain-size of the Table is also arbitrary. For example, both vividness and affect substitution (e.g. Slovic et al., 2002) could be viewed as simply specific aspects of the general phenomenon of attribute substitution discussed by Kahneman and Frederick (2002, 2005).

Some tasks are cognitively complex, and it is no surprise that some of the most complex tasks are those that are among the most contentious in the heuristics and biases literature. Thus, the taxonomy argues indirectly that non-normative responding on some of these tasks is overdetermined. For example, conjunction errors on tasks such as the Linda problem could result from attribute substitution in the manner that Tversky and Kahneman (1983) originally argued, from conversational defaults of the type discussed by a host of theorists (Adler, 1984, 1991; Girotto, 2004; Mellers, Hertwig, and Kahneman, 2001; Politzer and Macchi, 2000), and/or such errors could be exacerbated by missing mindware—that is, inadequately instantiated probabilistic mindware that impairs not just probabilistic calculations but also the tendency to see a problem in probabilistic terms.

It is likewise with the selection task. The Figure illustrates the conjecture that focal bias of the type exemplified in the emphasis on the matching bias (Evans, 1972, 1998; Evans and Lynch, 1973) is implicated in selection task performance, as are the interpretational defaults emphasized by many theorists (Margolis, 1987; Oaksford and Chater, 1994; Osman and Laming, 2001). But the Figure also captures the way the task was often treated in the earliest years of research on it—as a proxy for Popperian falsifiability tendencies. From this latter standpoint, problems in dealing with the task might be analyzed as a missing mindware problem. Certainly training programs in the critical thinking literature consider the generation of alternative hypotheses and falsification strategies as learnable mindware (Nickerson, 2004; Nisbett, 1993; Perkins, 1995).

Of course, Figure 3.4 is not meant to be exhaustive. It is very much a preliminary sketch and it is not meant to be the final word on many definitional/conceptual issues. The taxonomy is meant to serve an organizing function, to provoke research on the conjectures implicit within it, and to demonstrate how a framework deriving from dual-process theory (as I conceive it) might bring some order to the unwieldy heuristics and biases literature.

Conclusion

In summary, what has been termed System 2 in the dual-process literature is composed of (at least) the distinct operations of the reflective and algorithmic level of analysis. Rationality is a function of processes at both the reflective and algorithmic levels—

Table 3.1 A basic taxonomy of rational thinking errors

	The Cognitive Miser			Mindware Gaps (MG)		MG & (CM)	Contaminated Mindware (CM)	
Tasks, Effects, and Processing Styles	**Default to the Autonomous Mind**	**Focal Bias**	**Override Failure**	**Probability Knowledge**	**Alternative Thinking**	**Lay Psychological Theory**	**Evaluation Disabling Strategies**	**Self and Egocentric Processing**
Vividness effects	X							
Affect substitution	X							
Impulsively associative thinking	X							
Framing effects		X						
Anchoring effects		X						
Belief bias			X					
Denominator neglect			X					
Outcome bias			X					
Hindsight bias ('curse of knowledge' effects)			X					
Self-control problems			X					
Noncausal baserates				X				

Gambler's fallacy				X				
Bias blind spot						X		
Noncausal baserates				X				
Causal baserates			X	X				
Conjunction errors	X			X				
Ignoring P(D/~H)		X			X			
Four card selection task	X	X			X			
Myside processing		X						X
Affective forecasting errors			X				X	
Confirmation bias		X			X		X	
Overconfidence effects		X						X
Probability matching		X		X				
Pseudoscientific beliefs					X		X	
Evaluability effects		X				X		

specifically, thinking dispositions and fluid intelligence. In terms of human individual differences, rationality is thus a more encompassing construct than intelligence. Theoretical and empirical evidence for a tripartite structure was presented. Empirically, thinking dispositions and fluid intelligence predict unique portions of variance in performance on heuristics and biases tasks. Theoretically, it was argued that it makes sense to distinguish the cognitive control signal to begin decoupling operations from the ability to sustain decoupled representations. A distinct type of System 2 processing—serial associative cognition—was also defined. The importance of the knowledge bases recruited by each system in the tripartite model was stressed. All of these insights were fused into a taxonomy for classifying the thinking problems that people have on heuristics and biases tasks.

References

Ackerman, P. L., and Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, **121**, 219–45.

Ackerman, P. L., and Kanfer, R. (2004). Cognitive, affective, and conative aspects of adult intellect within a typical and maximal performance framework. In D. Y. Dai and R. J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (119–41). Mahwah, NJ: Lawrence Erlbaum Associates.

Adler, J. E. (1984). Abstraction is uncooperative. *Journal for the Theory of Social Behaviour*, **14**, 165–81.

Adler, J. E. (1991). An optimist's pessimism: Conversation and conjunctions. In E. Eells and T. Maruszewski (Eds.), *Probability and rationality: Studies on L. Jonathan Cohen's philosophy of science* (251–82). Amsterdam: Editions Rodopi.

Ainslie, G. (2001). *Breakdown of will*. Cambridge: Cambridge University Press.

Anderson, M. (1998). Mental retardation, general intelligence, and modularity. *Learning and Individual Differences*, **10**, 159–78.

Arkes, H. R., and Ayton, P. (1999). The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological Bulletin*, **125**, 591–600.

Aunger, R. (Ed.). (2000). *Darwinizing culture: The status of memetics as a science*. Oxford: Oxford University Press.

Baddeley, A., Chincotta, D., and Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, **130**, 641–57.

Bechara, A. (2005). Decision making, impulse control and loss of willpower to resist drugs: A neurocognitive perspective. *Nature Neuroscience*, **8**, 1458–63.

Bechara, A., Damasio, A. R., Damasio, H., and Anderson, S. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, **50**, 7–15.

Bermudez, J. L. (2001). Normativity and rationality in delusional psychiatric disorders. *Mind & Language*, **16**, 457–93.

Blackmore, S. (1999). *The meme machine*. New York: Oxford University Press.

Blackmore, S. (2005). Can memes meet the challenge? In S. Hurley and N. Chater (Eds.), *Perspectives on imitation* (Vol. 2, pp.409–11). Cambridge: MIT Press.

Brase, G. L. (2004). What we reason about and why: How evolution explains reasoning. In K. I. Manktelow and M. C. Chung (Eds.), *Psychology of reasoning: Theoretical and historical perspectives* (309–31). Hove, England: Psychology Press.

Bruine de Bruin, W., Parker, A. M., and Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology*, **92**, 938–56.

Buckner, R. L., and Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, **11**, 49–57.

Byrne, R. M. J. (2005). The rational imagination: How people create alternatives to reality. Cambridge, MA: MIT Press.

Cacioppo, J. T., Petty, R. E., Feinstein, J., and Jarvis, W. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, **119**, 197–253.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, **25**, 657–726.

Carruthers, P. (2006). *The architecture of the mind*. New York: Oxford University Press.

Churchland, P. M. (1988). *Matter and consciousnes - Revised edition*. Cambridge, MA: MIT Press.

Clark, A. (2001). Mindware: An introduction to the philosophy of cognitive science. New York: Oxford University Press.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, **4**, 317–70.

Conway, A. R. A., Kane, M. J., and Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Science*, **7**, 547–52.

Cosmides, L., and Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, and J. Tooby (Eds.), *The adapted mind* (163–228). New York: Oxford University Press.

Cosmides, L., and Tooby, J. (2000). Consider the source: The evolution of adaptations for decoupling and metarepresentation. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (53–115). Oxford: Oxford University Press.

Damasio, A. R. (1994). *Descartes' error*. New York: Putnam.

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society (London)*, **351**, 1413–20.

Davies, M. (2000). Interaction without reduction: The relationship between personal and sub-personal levels of description. *Mind & Society*, **1**, 87–105.

Dawes, R. M. (1976). Shallow psychology. In J. S. Carroll and J. W. Payne (Eds.), *Cognition and social behavior* (3–11). Hillsdale, NJ: Erlbaum.

De Neys, W. (2006a). Automatic-heuristic and executive-analytic processing during reasoning: Chronometric and dual-task considerations. *Quarterly Journal of Experimental Psychology*, **59**, 1070–1100.

De Neys, W. (2006b). Dual processing in reasoning—Two systems but one reasoner. *Psychological Science*, **17**, 428–33.

Dennett, D. C. (1984). *Elbow room: The varieties of free will worth wanting*. Cambridge, MA: MIT Press.

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

Dennett, D. C. (1995). Darwin's dangerous idea: Evolution and the meanings of life. New York: Simon & Schuster.

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

Dennett, D. C. (2006). From typo to thinko: When evolution graduated to semantic norms. In S. C. Levinson and P. Jaisson (Eds.), *Evolution and culture* (133–45). Cambridge, MA: MIT Press.

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

Distin, K. (2005). *The selfish meme*. Cambridge: Cambridge University Press.

Duckworth, A. L., and Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, **16**, 939–44.

Duncan, J., Emslie, H.,Williams, P., Johnson, R., and Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology*, **30**, 257–303.

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., Newell, F. N., and Emslie, H. (2000). A neural basis for general intelligence. *Science*, **289**, 457–60.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, **11**, 19–23.

Eslinger, P. J., and Damasio, A.R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology*, **35**, 1731–41.

Evans, J. St. B. T. (1972). Interpretation and matching bias in a reasoning task. *Quarterly Journal of Experimental Psychology*, **24**, 193–9.

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, J. St. B. T. (1998). Matching bias in conditional reasoning: Do we understand it after 25 years? *Thinking & Reasoning*, **4**, 45–82.

Evans, J. St. B. T. (2002). The influence of prior belief on scientific thinking. In P. Carruthers, S. Stich, and M. Siegal (Eds.), *The cognitive basis of science* (193–210). Cambridge: Cambridge University Press.

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, **7**, 454–9.

Evans, J. St. B. T. (2006a). Dual system theories of cognition: Some issues. *Proceedings of the 28th Annual Meeting of the Cognitive Science Society, Vancouver*, 202–7.

Evans, J. St. B. T. (2006b). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin and Review*, **13**, 378–95.

Evans, J. S. B. T. (2007). Hypothetical thinking: Dual processes in reasoning and judgment. New York: Psychology Press.

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

Evans, J. St. B. T., and Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, **11**, 382–9.

Evans, J. St. B. T., and Feeney, A. (2004). The role of prior belief in reasoning. In J. P. Leighton and R. J. Sternberg (Eds.), *The nature of reasoning* (78–102). Cambridge: Cambridge University Press.

Evans, J. St. B. T., and Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology*, **64**, 391–7.

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

Evans, J. St. B. T., and Over, D. E. (1999). Explicit representations in hypothetical thinking. *Behavioral and Brain Sciences*, **22**, 763–4.

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

Evans, J. St. B. T., Over, D. E., and Handley, S. J. (2003). A theory of hypothetical thinking. In D. Hardman and L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgment and decision making* New York: John Wiley.

Evans, J. St. B. T., and Wason, P. C. (1976). Rationalization in a reasoning task. *British Journal of Psychology*, **67**, 479–86.

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

Franklin, S. (1995). *Artificial Minds*. Cambridge, MA: MIT Press.

Gauthier, D. (1986). *Morals by agreement*. Oxford: Oxford University Press.

Geary, D. C. (2005). The origin of the mind: Evolution of brain, cognition, and general intelligence. Washington, DC: American Psychological Association.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review*, **103**, 592–6.

Gilbert, D. (1991). How mental systems believe. *American Psychologist*, **46**, 107–19.

Gilbert, D. (2006). *Stumbling on happiness*. New York: Alfred A. Knopf.

Girotto, V. (2004). Task understanding. In J. P. Leighton and R. J. Sternberg (Eds.), *The nature of reasoning* (103–25). Cambridge: Cambridge University Press.

Gould, S. J. (1991). *Bully for the Brontosaurus*. New York: Norton.

Hasher, L., Lustig, C., and Zacks, R. (2007). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake, and J. Towse (Eds.), *Variation in working memory* (227–49). New York: Oxford University Press.

Hasher, L., Zacks, R. T., and May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher and A. Koriat (Eds.), *Attention and Performance XVII, Cognitive Regulation of Performance: Interaction of Theory and Application* (653–75). Cambridge, MA: MIT Press.

Heaton, R., Chelune, G., Talley, J., Kay, G., and Curtiss, G. (1993). *Wisconsin Card Sorting Test - Revised and expanded*. Lutz, FL: Psychological Assessment Resource.

Horn, J. L., and Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. Flanagan, J. Genshaft, and P. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (53–91). New York: Guilford Press.

Hsee, C. K., and Hastie, R. (2006). Decision and experience: Why don't we choose what makes us happy? *Trends in Cognitive Sciences*, **10**, 31–7.

Hull, D. L. (2000). Taking memetics seriously: Memetics will be what we make it. In R. Aunger (Ed.), *Darwinizing culture: The status of memetics as a science* (43–67). Oxford: Oxford University Press.

Hull, D. L. (2001). Science and selection: Essays on biological evolution and the philosophy of science. Cambridge: Cambridge University Press.

Jackendoff, R. (1996). How language helps us think. *Pragmatics and Cognition*, **4**, 1–34.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology*, **50**, 109–35.

Johnson-Laird, P. N. (2005). Mental models and thought. In K. J. Holyoak and R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (185–208). New York: Cambridge University Press.

Kagan, J., Rosman, B. L., Day, D., Albert, J., and Philips, W. (1964). Information processing in the child: Significance of analytic and reflective attitudes. *Psychological Monographs*, **78**, 578.

Kahneman, D. (2000). A psychological point of view: Violations of rational rules as a diagnostic of mental processes. *Behavioral and Brain Sciences*, **23**, 681–3.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, **58**, 697–720.

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

Kahneman, D., and Frederick, S. (2005). A model of heuristic judgment. In K. J. Holyoak and R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (267–93). New York: Cambridge University Press.

Kahneman, D., and Tversky, A. (1982a). On the study of statistical intuitions. *Cognition*, **11**, 123–41.

Kahneman, D., and Tversky, A. (1982b). The simulation heuristic. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (201–8). Cambridge: Cambridge University Press.

Kahneman, D., and Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, **103**, 582–91.

Kahneman, D., and Tversky, A. (Eds.). (2000). *Choices, values, and frames*. Cambridge: Cambridge University Press.

Kane, M. J. (2003). The intelligent brain in conflict. *Trends in Cognitive Sciences*, **7**, 375–7.

Kane, M. J., Bleckley, M., Conway, A., and Engle, R. W. (2001). A controlled-attention view of WM capacity. *Journal of Experimental Psychology: General*, **130**, 169–83.

Kane, M. J., and Engle, R. W. (2002). The role of prefrontal cortex working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin and Review*, **9**, 637–71.

Kane, M. J., and Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, **132**, 47–70.

Kane, M. J., Hambrick, D. Z., and Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, **131**, 66–71.

Keren, G. (1994). The rationality of gambling: Gamblers' conceptions of probability, chance and luck. In G. Wright and P. Ayton (Eds.), *Subjective probability* (485–99). Chichester, UK: Wiley.

Kimberg, D. Y., D'Esposito, M., and Farah, M. J. (1998). Cognitive functions in the prefrontal cortex—working memory and executive control. *Current Directions in Psychological Science*, **6**, 185–92.

Klaczynski, P. A., and Lavallee, K. L. (2005). Domain-specific identity, epistemic regulation, and intellectual ability as predictors of belief-based reasoning: A dual-process perspective. *Journal of Experimental Child Psychology*.

Kokis, J, Macpherson, R., Toplak, M., West, R. F., and Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, **83**, 26–52.

Krueger, J., and Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences and cures for the problem-seeking approach to social cognition and behavior. *Behavioral and Brain Sciences*, **27**, 313–76.

Lee, C. J. (2006). Gricean charity: The Gricean turn in psychology. *Philosophy of the Social Sciences*, **36**, 193–218.

Legrenzi, P., Girotto, V., and Johnson-Laird, P. N. (1993). Focussing in reasoning and decision making. *Cognition*, **49**, 37–66.

Lepine, R., Barrouillet, P., and Camos, V. (2005). What makes working memory spans so predictive of high-level cognition? *Psychonomic Bulletin & Review*, **12**, 165–70.

Leslie, A. M. (1987). Pretense and representation: The origins of 'Theory of Mind'. *Psychological Review*, **94**, 412–26.

Lucas, E. J., and Ball, L. J. (2005). Think-aloud protocols and the selection task: Evidence for relevance effects and rationalisation processes. *Thinking & Reasoning*, **11**, 35–66.

Lynch, A. (1996). *Thought contagion*. New York: Basic Books.

Manktelow, K. I. (2004). Reasoning and rationality: The pure and the practical. In K. I. Manktelow and M. C. Chung (Eds.), *Psychology of reasoning: Theoretical and historical perspectives* (157–77). Hove, England: Psychology Press.

Margolis, H. (1987). *Patterns, thinking, and cognition*. Chicago: University of Chicago Press.

Markovits, H., and Nantel, G. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition*, **17**, 11–17.

McCarthy, R. A., and Warrington, E. K. (1990). *Cognitive neuropsychology: A clinical introduction*. San Diego: Academic Press.

Mellers, B., Hertwig, R., and Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science*, **12**, 269–75.

Mesoudi, A.,Whiten, A., and Laland, K. N. (2006). Towards a unified science of cultural evolution. *Behavioral and Brain Sciences*, **29**, 329–83.

Mithen, S. (1996). The prehistory of mind: The cognitive origins of art and science. London: Thames and Hudson.

Miyake, A., Friedman, N., Emerson, M. J., and Witzki, A. H. (2000). The utility and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: A latent variable analysis. *Cognitive Psychology,* **41**, 49–100.

Murphy, D., and Stich, S. (2000). Darwin in the madhouse: Evolutionary psychology and the classification of mental disorders. In P. Carruthers and A. Chamberlain (Eds.), *Evolution and the human mind: Modularity, language and meta-cognition* (62–92). Cambridge: Cambridge University Press.

Nichols, S., and Stich, S. P. (2003). Mindreading: An integrated account of pretence, self-awareness, and understanding other minds. Oxford: Oxford University Press.

Nickerson, R. S. (2004). Teaching reasoning. In J. P. Leighton and R. J. Sternberg (Eds.), *The nature of reasoning* (410–42). Cambridge: Cambridge University Press.

Nisbett, R. E. (1993). *Rules for reasoning*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Oaksford, M., and Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, **101**, 608–31.

Oatley, K. (1999). Why fiction may be twice as true as fact: Fiction as cognitive and emotional simulation. *Review of General Psychology*, **3**, 101–17.

Osman, M., and Laming, D. (2001). Misinterpretation of conditional statements in Wason's selection task. *Psychological Research*, **65**, 128–44.

Over, D. E. (2002). The rationality of evolutionary psychology. In J. L. Bermudez and A. Millar (Eds.), *Reason and nature: Essays in the theory of rationality* (187–207). Oxford: Oxford University Press.

Over, D. E. (2004). Rationality and the normative/descriptive distinction. In D. J. Koehler and N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (3–18). Malden, MA: Blackwell Publishing.

Parker, A. M., and Fischhoff, B. (2005). Decision-making competence: External validation through an individual differences approach. *Journal of Behavioral Decision Making*, **18**, 1–27.

Pennington, B. F., and Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, **37**, 51–87.

Perkins, D. N. (1995). Outsmarting IQ: The emerging science of learnable intelligence. New York: Free Press.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Politzer, G., and Macchi, L. (2000). Reasoning and pragmatics. *Mind & Society*, **1**, 73–93.

Pollock, J. L. (1995). Cognitive carpentry: A blueprint for how to build a person. Cambridge, MA: MIT Press.

Pronin, E. (2006). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences*, **11**, 37–43.

Roberts, M. J., and Newton, E. J. (2001). Inspection times, the change task, and the rapid-response selection task. *Quarterly Journal of Experimental Psychology*, **54A**, 1031–48.

Rokeach, M. (1960). *The open and closed mind*. New York: Basic Books.

Sá, W., Kelley, C., Ho, C., and Stanovich, K. E. (2005). Thinking about personal theories: Individual differences in the coordination of theory and evidence. *Personality and Individual Differences*, **38**, 1149–61.

Sá, W., West, R. F., and Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, **91**, 497–510.

Salthouse, T. A., Atkinson, T. M., and Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*, **132**, 566–94.

Samuels, R. (2005). The complexity of cognition: Tractability arguments for massive modularity. In P. Carruthers, S. Laurence, and S. Stich (Eds.), *The innate mind* (107–121). Oxford: Oxford University Press.

Samuels, R., and Stich, S. P. (2004). Rationality and psychology. In A. R. Mele and P. Rawling (Eds.), *The Oxford handbook of rationality* (279–300). Oxford: Oxford University Press.

Samuels, R., Stich, S. P., and Bishop, M. (2002). Ending the rationality wars: How to make disputes about human rationality disappear. In R. Elio (Ed.), *Common sense, reasoning and rationality* (236–68). New York: Oxford University Press.

Samuels, R., Stich, S. P., and Tremoulet, P. D. (1999). Rethinking rationality: From bleak implications to Darwinian modules. In E. Lepore and Z. Pylyshyn (Eds.), *What is cognitive science?* (74–120). Oxford: Blackwell.

Sanfey, A. G., Loewenstein, G., McClure, S. M., and Cohen, J. D. (2006). Neuroeconomics: Cross-currents in research on decision-making. *Trends in Cognitive Sciences*, **10**, 108–16.

Shafir, E. (1994). Uncertainty and the difficulty of thinking through disjunctions. *Cognition*, **50**, 403–30.

Shafir, E., and LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology*, **53**, 491–517.

Shiffrin, R. M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, **84**, 127–90.

Sloman, A. (1993). The mind as a control system. In C. Hookway and D. Peterson (Eds.), *Philosophy and cognitive science* (69–110). Cambridge: Cambridge University Press.

Sloman, A., and Chrisley, R. (2003). Virtual machines and consciousness. *Journal of Consciousness Studies*, **10**, 133–72.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, **119**, 3–22.

Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (379–96). New York: Cambridge University Press.

Slovic, P., Finucane, M. L., Peters, E., and MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (397–420). New York: Cambridge University Press.

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld and S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (39–67). Cambridge: Cambridge University Press.

Sperber, D. (2000). Metarepresentations in evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective* (117–37). Oxford: Oxford University Press.

Sperber, D., Cara, F., and Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, **57**, 31–95.

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

Stanovich, K. E. (2001). Reductionism in the study of intelligence: Review of 'Looking Down on Human Intelligence' by Ian Deary. *Trends in Cognitive Sciences*, **5**(2), 91–2.

Stanovich, K. E. (2002). Rationality, intelligence, and levels of analysis in cognitive science: Is dysrationalia possible? In R. J. Sternberg (Ed.), *Why smart people can be so stupid* (124–58). New Haven, CT: Yale University Press.

Stanovich, K. E. (2003). The fundamental computational biases of human cognition: Heuristics that (sometimes) impair decision making and problem solving. In J. E. Davidson and R. J. Sternberg (Eds.), *The psychology of problem solving* (291–342). New York: Cambridge University Press.

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

Stanovich, K. E. (in press a). Rationality and the reflective mind: Toward a tri-process model of cognition. New York: Oxford University Press.

Stanovich, K. E. (in press b). What IQ tests miss: The cognitive science of rational and irrational thinking. New Haven, CT: Yale University Press.

Stanovich, K. E., and West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, **89**, 342–57.

Stanovich, K. E., and West, R. F. (1998a). Cognitive ability and variation in selection task performance. *Thinking & Reasoning*, **4**, 193–230.

Stanovich, K. E., and West, R. F. (1998b). Individual differences in framing and conjunction effects. *Thinking & Reasoning*, **4**, 289–317.

Stanovich, K. E., and West, R. F. (1998c). Individual differences in rational thought. *Journal of Experimental Psychology: General*, **127**, 161–88.

Stanovich, K. E., and West, R. F. (1998d). Who uses base rates and P(D/~H)? An analysis of individual differences. *Memory & Cognition*, **26**, 161–79.

Stanovich, K. E., and West, R. F. (1999). Discrepancies between normative and descriptive models of decision making and the understanding/acceptance principle. *Cognitive Psychology*, **38**, 349–85.

Stanovich, K. E., and West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, **23**, 645–726.

Stanovich, K. E., and West, R. F. (2003). Evolutionary versus instrumental goals: How evolutionary psychology misconceives human rationality. In D. Over (Ed.), *Evolution and the psychology of thinking: The debate* (171–230). Hove, England: Psychology Press.

Stanovich, K. E., and West, R. F. (2007). Natural my side bias is independent of cognitive ability. *Thinking & Reasoning*, **13**, 225–47.

Stanovich, K. E., and West, R. F. (in press). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*.

Stein, E. (1996). Without good reason: The rationality debate in philosophy and cognitive science. Oxford: Oxford University Press.

Suddendorf, T., and Whiten, A. (2001). Mental evolution and development: Evidence for secondary representation in children, great apes, and other animals. *Psychological Bulletin*, **127**, 629–50.

Taylor, S. E. (1981). The interface of cognitive and social psychology. In J. H. Harvey (Ed.), *Cognition, social behavior, and the environment* (189–211). Hillsdale, NJ: Erlbaum.

Toplak, M., Liu, E., Macpherson, R., Toneatto, T., and Stanovich, K. E. (2007). The reasoning skills and thinking dispositions of problem gamblers: A dual-process taxonomy. *Journal of Behavioral Decision Making*, **20**, 103–24.

Toplak, M., and Stanovich, K. E. (2002). The domain specificity and generality of disjunctive reasoning: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, **94**, 197–209.

Toplak, M. E. and Stanovich, K. E. (2003). Associations between my side bias on an informal reasoning task and amount of post-secondary education. *Applied Cognitive Psychology*, **17**, 851–60.

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 1124–31.

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

Tweney, R. D., and Yachanin, S. (1985). Can scientists rationally assess conditional inferences? *Social Studies of Science*, **15**, 155–73.

Unsworth, N., and Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence*, **33**, 67–81.

Wagenaar, W. A. (1988). *Paradoxes of gambling behavior*. Hove, England: LEA.

Wason, P. C., and Evans, J. St. B. T. (1975). Dual processes in reasoning? *Cognition*, **3**, 141–54.

Watson, G., and Glaser, E. M. (1980). *Watson-Glaser Critical Thinking Appraisal*. New York: Psychological Corporation.

West, R. F., and Stanovich, K. E. (2003). Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory & Cognition*, **31**, 243–51.

Whiten, A. (2001). Meta-representation and secondary representation. *Trends in Cognitive Sciences*, **5**, 378.

Wilson, T. D., Wheatley, T., Meyers, J. M.,Gilbert, D. T., and Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, **78**, 821–36.