CHAPTER

22

On the Distinction Between Rationality and Intelligence: Implications for Understanding Individual Differences in Reasoning

Keith E. Stanovich

Abstract

A concern for individual differences has been missing from the Great Rationality Debate in cognitive science—the debate about how much irrationality to attribute to human cognition. There are individual differences in rational thinking that are less than perfectly correlated with individual differences in intelligence because intelligence and rationality occupy different conceptual locations in models of cognition. A tripartite extension of currently popular dual-process theories is presented in this chapter that illustrates how intelligence and rationality are theoretically separate concepts. The chapter concludes by showing how this tripartite model of mind, taken in the context of studies of individual differences, can help to resolve the Great Rationality Debate.

Key Words: rationality, intelligence, reasoning, individual differences

Introduction

In psychology and among the lay public alike, assessments of intelligence and tests of cognitive ability are taken to be the sine qua non of good thinking. Critics of these instruments often point out that IQ tests fail to assess many domains of psychological functioning that are essential. For example, many largely noncognitive domains such as socioemotional abilities, creativity (Smith & Ward, Chapter 23), empathy, and interpersonal skills are almost entirely unassessed by tests of cognitive ability. However, even these common critiques of intelligence tests often contain the unstated assumption that although intelligence tests miss certain key noncognitive areas, they encompass most of what is important cognitively. In this chapter, I will challenge this tacit assumption by arguing that certain very important classes of individual differences in thinking are ignored if only intelligence-related variance is the focus. Many of these classes of individual differences that are missing from IQ tests are those relating to rational thought (Chater & Oaksford, Chapter 2). In this chapter, I will illustrate how a comprehensive assessment of individual differences in reasoning skills will necessitate the theoretical and empirical differentiation of the concepts of intelligence and rational thinking.

The discussion in this chapter will begin by showing how differing definitions of rationality frame what is known as the Great Rationality Debate in cognitive science. That debate concerns how much irrationality to attribute to human cognition. I will argue that a concern for individual differences has been missing from this debate because we failed to appreciate that there are individual differences in rational thought as well as intelligence. This is easier to appreciate when we realize that intelligence and rationality occupy somewhat different conceptual locations within most models of cognition. Thus, I present a generic model of the mind that is an extension of currently popular dual-process theories (Evans, Chapter 8), and I situate both intelligence and rationality within this model and show how they dissociate, both conceptually and empirically.

I conclude the chapter by showing how this tripartite model of mind, taking in the context of studies of individual differences, can help to resolve the Great Rationality Debate.

The Concept of Rational Thought in Cognitive Science and Philosophy

The term *rationality* has a strong and a weak sense. The strong sense of the term is the one used in cognitive science, and it will be the one used throughout this chapter. However, a weaker sense of the term has sometimes influenced—and hence confused—arguments in the so-called Great Rationality Debate in cognitive science. The influence of the weak sense of the term has also impeded investigation into individual differences in rational thought.

Dictionary definitions of rationality tend to be of the weak sort—often seeming quite lame and unspecific ("the state or quality of being in accord with reason"). The meaning of rationality in modern cognitive science (the strong sense) is, in contrast, much more specific and prescriptive than this. The weak definitions of rationality derive from a categorical notion of rationality tracing to Aristotle (humans as the only animals who base actions on reason). As de Sousa (2007) has pointed out, such a notion of rationality as "based on reason" has as its opposite not irrationality but *arationality* (outside the domain of reason). Aristotle's characterization is categorical—the behavior of entities is either based on thought or it is not. Animals are either rational or arational. In this conception, humans are rational, but other animals are not. There is no room for individual differences in rational thinking *among* humans in this view.

In its stronger sense (the sense employed in most of cognitive science and in this chapter) rational thought is a normative notion (Chater & Oaksford, Chapter 2; Griffiths, Tenenbaum, & Kemp, Chapter 3). Its opposite is irrationality, not arationality. Normative models of optimal judgment and decision making define perfect rationality in the noncategorical view employed in cognitive science. Rationality (and irrationality) comes in degrees defined by the distance of the thought or behavior from the optimum defined by a normative model. De Sousa (2007) points out that the notion of rationality in Aristotle's sense cannot be normative, but in the strong sense of cognitive science, it is. Other animals may be arational, but only humans can be irrational. As de Sousa (2007) puts it, "if human beings can indeed be described as rational animals, it is precisely in virtue of the fact that humans, of all the animals, are the only ones capable of irrational thoughts and actions" (p. 7).

Hurley and Nudds (2006) make a similar point when they argue that, for a strong sense of the term: "ironically, rationality requires the possibility that the animal might err. It can't be automatically right, no matter what it does. . . . when we say that an agent has acted rationally, we imply that it would have been a mistake in some sense to have acted in certain different ways. It can't be the case that anything the agent might do would count as rational. This is normativity in a quite weak sense" (p. 2). The weak sense they are referring to is an Aristotelian (categorical) sense, and no cognitive scientist is using rationality in this sense when claiming that an experiment has demonstrated human irrationality.

When a cognitive scientist terms a behavior irrational, he or she means that the behavior departs from the optimum prescribed by a particular normative model. The scientist is not implying that no thought or reason was behind the behavior. Some of the hostility that has been engendered by experimental claims of human irrationality no doubt derive from a (perhaps tacit) influence of the Aristotelian view—the thought that cognitive psychologists are saying that certain people are somehow less than human when they are said to behave irrationally. But in using the strong sense of the term *rationality*, most cognitive scientists are saying no such thing.[1]

Some of the heat in the Great Rationality Debate is no doubt caused by reactions to the term *irrationality* being applied to humans. As mentioned, lingering associations with the Aristotelian categorical view make charges of irrationality sound more cutting than they actually are when in the context of cognitive science research. When we find a behavioral pattern that is less than optimally rational, we could easily say that it is "less than perfectly rational" rather than that it is irrational—with no loss of meaning. Perhaps if this had been the habit in the literature, the rationality debate in cognitive science would not have become so heated. Such an emphasis also highlights the theme of this chapter—that there are indeed individual differences in rational thought and that understanding the nature of these differences might have important theoretical implications.

Cognitive scientists recognize two types of rationality: epistemic and instrumental. *Epistemic rationality* concerns how well beliefs map onto the actual structure of the world. It is sometimes called theoretical rationality or evidential rationality (see

Audi, 1993, 2001; Foley, 1987; Harman, 1995; Manktelow, 2004; Over, 2004).

The simplest definition of *instrumental rationality* is as follows: behaving in the world so that you get exactly what you most want, given the resources (physical and mental) available to you. Somewhat more technically, we could characterize instrumental rationality as the optimization of the individual's goal fulfillment. Economists and cognitive scientists have refined the notion of optimization of goal fulfillment into the technical notion of expected utility. The model of rational judgment used by decision scientists (Chater & Oaksford, Chapter 2; LeBoeuf & Shafir, Chapter 16) is one in which a person chooses options based on which option has the largest expected utility[2] (see Baron, 2008; Dawes, 1998; Hastie & Dawes, 2010; Wu, Zhang, & Gonzalez, 2004). One of the fundamental advances in the history of modern decision science was the demonstration that if people's preferences follow certain patterns (the so-called axioms of choice—things like transitivity and freedom from certain kinds of context effects), then they are behaving as if they are maximizing utility; they are acting to get what they most want (Edwards, 1954; Gilboa, 2010; Jeffrey, 1983; Luce & Raiffa, 1957; Savage, 1954; von Neumann & Morgenstern, 1944). This is what makes people's degrees of rationality measurable by the experimental methods of cognitive science. Although it is difficult to assess utility directly, it is much easier to assess whether one of the axioms of rational choice is being violated. This is much like our judgments at a sporting event, where, for example, it might be difficult to discern whether a quarterback has put the ball perfectly on the money, but it is not difficult at all to detect a bad throw.

In fact, in many domains of life this is often the case as well. It is often difficult to specify what the very *best* response might be, but performance *errors* are much easier to spot. Essayist Neil Postman (1988) has argued, for instance, that educators and other advocates of good thinking might adopt a stance more similar to that of physicians or attorneys. He points out that doctors would find it hard to define "perfect health" but, despite this, they are quite good at spotting disease. Likewise, lawyers are much better at spotting injustice and lack of citizenship than defining "perfect justice" or ideal citizenship. Postman argues that, like physicians and attorneys, educators might best focus on instances of poor thinking which are much easier to identify as opposed to trying to define ideal thinking. The literature on the psychology of rationality has followed this logic in that the empirical literature has focused on identifying thinking errors, just as physicians focus on disease. Degrees of rationality can be assessed in terms of the number and severity of such cognitive biases that individuals display. Conversely, *failure* to display a cognitive bias becomes a measure of rational thought.

The Great Rationality Debate in Cognitive Science

A substantial research literature—one comprising literally hundreds of empirical studies conducted over several decades—has firmly established that people's responses sometimes deviate from the performance considered normative on many reasoning tasks. For example, people assess probabilities incorrectly, they test hypotheses inefficiently, they violate the axioms of utility theory, they do not properly calibrate degrees of belief, their choices are affected by irrelevant context, they ignore the alternative hypothesis when evaluating data, and they display numerous other information processing biases (Baron, 2008; Bazerman & Moore, 2008; Evans, 2007; Gilovich, Griffin, & Kahneman, 2002; Kahneman & Tversky, 2000; Shafir & LeBoeuf, 2002; Stanovich, 2009b, 2011). Demonstrating that descriptive accounts of human behavior diverged from normative models was a main theme of the heuristics and biases research program inaugurated by Kahneman and Tversky in the early 1970s (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1974).

Researchers working in the heuristics and biases tradition tend to be so-called Meliorists (see Bishop & Trout, 2005; Doherty, 2003; Larrick, 2004; Stanovich, 1999, 2004). They assume that human reasoning is not as good as it could be, and that thinking could be improved. The dictionary definition of *meliorism* is "the doctrine that the world tends to become better or may be made better by human effort." Thus, a Meliorist is one who feels that education and the provision of information could help make people more rational—could help them more efficiently further their goals and to bring their beliefs more in line with the actual state of the world.[3] Stated this way, Meliorism seems to be an optimistic doctrine, and in one sense it is. But this optimistic part of the Meliorist message derives from the fact that Meliorists see a large gap between normative models of rational responding and descriptive models of what people actually do. Emphasizing the gap, of course, entails that

Meliorists will be attributing a good deal of irrationality to human cognition.

Over the last two decades, an alternative interpretation of the findings from the heuristics and biases research program has been championed. Contributing to this alternative interpretation have been evolutionary psychologists, adaptationist modelers, and ecological theorists (Anderson, 1990; Cosmides & Tooby, 1996; Gigerenzer, 2007; Marewski, Gaissmaier, & Gigerenzer, 2010; Oaksford & Chater, 2007). They have reinterpreted the modal response in most of the classic heuristics and biases experiments as indicating an optimal information processing adaptation on the part of the subjects. It is argued by these investigators that the research in the heuristics and biases tradition has not demonstrated human irrationality at all. This group of theorists—who argue that an assumption of maximal human rationality is the proper default position to take—have been termed the Panglossians (Stanovich, 1999). This position posits no difference between descriptive and normative models of performance because human performance is actually normative.

The contrasting positions of the Panglossians and Meliorists define the differing poles in what has been termed the Great Rationality Debate in cognitive science—the debate about how much irrationality to attribute to human cognition. This debate has generated a very substantial literature of often heated arguments (Cohen, 1981; Doherty, 2003; Edwards & von Winterfeldt, 1986; Evans & Over, 1996, 2010; Gigerenzer, 1996; Jungermann, 1986; Kahneman & Tversky, 1983, 1996; Koehler, 1996; Koehler & James, 2009, 2010; Krueger & Funder, 2004; Kuhberger, 2002; Lee, 2006; Samuels & Stich, 2004; Stanovich, 1999, 2004, 2010; Stanovich & West, 2000; Stein, 1996; Stich, 1990; Vranas, 2000). Tetlock and Mellers (2002) have noted that "the debate over human rationality is a high-stakes controversy that mixes primordial political and psychological prejudices in combustible combinations" (p. 97). The great debate about human rationality is a "high-stakes controversy" because it involves nothing less than the models of human nature that underlie economics, moral philosophy, and the personal theories (folk theories) we use to understand the behavior of other humans. For example, a very influential part of the Panglossian camp is represented by the mainstream of the discipline of economics, which is notable for using strong rationality assumptions as fundamental tools.

Evolution Does Not Guarantee Human Rationality

An Aristotelian view of rationality has no room for individual differences in rational thinking *between* humans. In this view, humans are the unique animals who act based on reason. Thus, all humans are rational—and all are equally so. However, once we move from this view to the normative conception of rationality, we open up room for individual differences. The maximizing notions of rational thought and action in decision science potentially array individuals on a continuum based on the distance of their behavior from the normative model.

We might ask, however, whether—aside from holding an Aristotelian view—there is any other reason to be a Panglossian. One assumption that often draws people to a Panglossian view of human rationality is the thought that evolution would have guaranteed that our cognition is fully rational. This is a mistaken view. There are a number of reasons why evolution would not be expected to guarantee perfect human rationality. One reason is that rationality is defined in terms of maximization (for example, in the case of instrumental rationality, maximizing the expected utility of actions). In contrast to maximization, natural selection works on a "better than" principle. As Dawkins puts it, "Natural selection chooses the better of present available alternatives.... The animal that results is not the most perfect design conceivable, nor is it merely good enough to scrape by. It is the product of a historical sequence of changes, each one of which represented, at best, the better of the alternatives that happened to be around at the time" (p. 46, 1982). In short, the variation and selective retention logic of evolution "designs" (Dennett, 1995) for the reproductive advantage of one organism over the next, not for the optimality of any one characteristic (including rationality). It has been said that evolution should be described as the survival of the *fitter* rather than as the survival of the fittest.

Evolution proceeds to increase the reproductive fitness of genes, not to increase the rationality of humans (Stanovich, 2004; Stanovich & West, 2003). Increases in fitness do not always entail increases in rationality. Take, for example, the domain of beliefs. Beliefs need not always track the world with maximum accuracy in order for fitness to increase (Stich, 1990). Thus, evolution does not guarantee perfect epistemic rationality. For example, evolution might fail to select out epistemic mechanisms of high accuracy when they are costly in terms of organismic resources (for example, in terms of memory, energy, or attention). An additional reason that belief-forming mechanisms might not be maximally truth preserving is that:

> a very cautious, risk-aversive inferential strategy—one that leaps to the conclusion that danger is present on very slight evidence—will typically lead to false beliefs more often, and true ones less often, than a less hair-trigger one that waits for more evidence before rendering a judgment. Nonetheless, the unreliable, error-prone, risk-aversive strategy may well be favored by natural selection. For natural selection does not care about truth; it cares only about reproductive success. (p. 62, Stich, 1990)

It is likewise in the domain of goals and desires. As has become clear from recent research on the topic of affective forecasting, people are remarkably bad at making choices that make themselves happy (Gilbert, 2006; Kahneman et al., 2006; Wilson & Gilbert, 2005). This should be no surprise. The reason we have pleasure circuits in our brains is to encourage us to do things (survive and reproduce, help kin) that propagate our genes. The pleasure centers were not designed to maximize the amount of time we are happy.

The instrumental rationality of humans is not guaranteed by evolution for two further reasons. First, many genetic goals that have been lodged in our brain no longer serve our ends because the environment has changed (Richerson & Boyd, 2005). For example, thousands of years ago, humans needed as much fat as they could get in order to survive. More fat meant longer survival and because few humans survived beyond their reproductive years, longevity translated directly into more opportunities for gene replication. In short, our mechanisms for storing and utilizing energy evolved in times when fat preservation was efficacious. These mechanisms no longer serve the goals of people in our modern technological society where there is a McDonald's on practically every corner—the goals underlying these mechanisms have become detached from their evolutionary context.

Finally, rational standards for assessing human behavior are social and cultural products that are preserved and stored independently of the genes. The development of probability theory, concepts of empiricism, logic, and scientific thinking throughout the centuries have provided humans with conceptual tools to aid in the formation and revision of belief and in their reasoning about action (Chater & Oaksford, Chapter 2; Griffiths et al., Chapter 3; Dunbar & Klahr, Chapter 35). They represent the cultural achievements that foster greater human rationality (Thagard & Nisbett, 1983). As societies evolve, they produce more of the cultural tools of rationality and these tools become more widespread in the population. Thus, the cultural evolution of rational standards (Thagard & Nisbett, 1983) is apt to occur at a pace markedly faster than that of human evolution—providing ample opportunity for mental strategies of utility maximization to dissociate from local genetic fitness maximization. In summary, a consideration of our evolutionary history should not lead one to a Panglossian view of human rationality.

A reconciliation of the views of the Panglossians and Meliorists is possible, however, if we take three scientific steps. First, we must consider data patterns long ignored in the heuristics and biases literature—individual differences on rational thinking tasks. Second, we must understand the empirical patterns obtained through the lens of a modified dual-process theory (Evans, Chapter 8) and of evolutionary theory. Thirdly, we much distinguish the concepts of rationality and intelligence in cognitive theory. Subsequent sections of this chapter develop each of these points.

Individual Differences in the Great Rationality Debate

Dozens of empirical studies have shown that there are few tasks in the heuristics and biases literature where all untutored laypersons give the same response. What has largely been ignored is that—although the average person in the classic heuristics and biases experiments might well display an overconfidence effect, underutilize base rates, ignore P(D/~H), violate the axioms of utility theory, choose P and Q in the selection task, commit the conjunction fallacy, and so on—on each of these tasks, *some people give the standard normative response* (Bruine de Bruin, Parker, & Fischhoff, 2007; Cokely & Kelley, 2009; Del Missier, Mantyla, & Bruine de Bruin, 2010; Dohmen et al., 2009; Finucane & Gullion, 2010; Frederick, 2005; Klaczynski, 2001; Oechssler, Roider, & Schmitz, 2009; Stanovich & West, 1998b, 1998c, 1999, 2000, 2008b; West et al., 2008). What has been ignored in the Great Rationality Debate is individual differences. For example, in knowledge calibration studies, although the mean performance level of the entire sample may be represented by a calibration curve that indicates overconfidence, almost always some people do display near perfect calibration. Likewise, in probabilistic assessment, while the majority of subjects might well ignore the noncausal base-rate evidence, a minority of subjects often makes use of this information in exactly the way prescribed by Bayes' theorem. A few people even respond correctly on the notoriously difficult abstract selection task (Evans, Newstead, & Byrne, 1993; Stanovich & West, 1998a, 2008b).

In short, some people give the response traditionally considered normative, and others do not. There is variability in responding on all of these tasks. So when Panglossians and heuristics and biases researchers argue about the normative appropriateness of a particular response, whoever eventually prevails in the dispute—both sides have been guilty of glossing over individual differences. In short, it is incorrect to imply that people uniformly display a particular rational or irrational response pattern. A particular experiment might instead be said to show that the average person, or perhaps the modal person, displays optimal or suboptimal thinking. Other people, often a minority to be sure, display the opposite style of thinking.

In light of these empirical data, it is puzzling that Panglossians would presumably accept the existence of individual differences in intelligence but not rationality. This is possible, however, if intelligence and rationality are two different things conceptually. In the remainder of this chapter, I will show that the Panglossians are correct in one of their assumptions but incorrect in another. Conceptually, intelligence and rational thinking are indeed two different things. But—contra Panglossian assumptions—the latter as well as the former displays substantial individual differences.

Discussions of intelligence often go off the rails at the very beginning by failing to set the concept within a general context of cognitive functioning—thus inviting the default assumption that intelligence is the central feature of the mind. I will try to preclude this natural default by outlining a model of the mind and then placing intelligence within it. The generic models of the mind developed by cognitive scientists often give short shrift to a question that the public is intensely interested in: How and why do people *differ* from each other in their thinking? In an attempt to answer that question, I am going to present a gross model of the mind that is true to modern cognitive science but that emphasizes individual differences in ways that are somewhat new. The model builds on a current consensus view of cognition termed dual-process theory (see Evans, Chapter 8, for a more detailed discussion).

From Dual-Process Theory to a Tripartite Model of Mind

The idea that the brain is composed of many different subsystems (see Aunger & Curtis, 2008) has recurred in conceptualizations in many different disciplines—from the society of minds view in artificial intelligence (Minsky, 1985); to Freudian analogies (Ainslie, 1982); to discussions of the concept of multiple selves in philosophy, economics, and decision science (Ainslie, 2001; Schelling, 1984). In fact, the notion of many different systems in the brain is by no means new. Plato (1945) argued that "we may call that part of the soul whereby it reflects, rational; and the other, with which it feels hunger and thirst and is distracted by sexual passion and all the other desires, we will call irrational appetite, associated with pleasure in the replenishment of certain wants" (p. 137).

What *is* new, however, is that cognitive scientists are beginning to understand the biology and cognitive structure of these systems (Evans & Frankish, 2009; see Morrison & Knowlton, Chapter 6; Green & Dunbar, Chapter 7) and are beginning to posit some testable speculations about their evolutionary and experiential origins. I will build on the current consensus that the functioning of the brain can be characterized by two different types of cognition having somewhat different functions and different strengths and weaknesses. There is a wide variety of evidence that has converged on the conclusion that some type of dual-process notion is needed in a diverse set of specialty areas not limited to cognitive psychology, social psychology, naturalistic philosophy, decision theory, and clinical psychology (Evans, 2003, 2008, 2010; Frankish, 2004; Lieberman, 2007, 2009; Schneider & Chein, 2003; Sloman, 1996, 2002; Smith & Decoster, 2000; Stanovich, 1999). Evolutionary theorizing and neurophysiological work also have supported a dual-process conception (Camerer, Loewenstein, & Prelec, 2005; Frank, Cohen, & Sanfey, 2009; Lieberman, 2009; McClure, Laibson, Loewenstein, & Cohen, 2004; Prado & Noveck, 2007; Reber, 1993; Toates, 2005, 2006). In fact, a dual-process view was implicit within the early writings in the groundbreaking heuristics and biases research program. As Kahneman (2000) notes, "Tversky and I always thought of the heuristics and biases approach as a two-process theory" (p. 682).

Just how ubiquitous are dual-process models in psychology and related fields is illustrated in

Table 22.1, which lists a variety of such theories that have appeared during the last couple of decades. Some common terms for the dual processes are listed in Table 22.1. My purpose here is not to adjudicate the differences among these models. Instead, I will gloss over differences and instead start with a model that emphasizes the family resemblances. Evans (Chapter 8) provides a much more nuanced discussion.

The family resemblances extend to the names for the two classes of process. The terms *heuristic* and *analytic* are two of the oldest and most popular (see

Table 22.1 Some Alternative Terms for Type 1 and Type 2 Processing Used by Various Theorists

Theorist	Type 1	Type 2
Bargh & Chartrand (1999)	Automatic processing	Conscious processing
Bazerman, Tenbrunsel, & Wade-Benzoni, (1998)	Want self	Should self
Bickerton (1995)	Online thinking	Offline thinking
Brainerd & Reyna (2001)	Gist processing	Analytic processing
Chaiken et al. (1989)	Heuristic processing	Systematic processing
Evans (1984, 1989)	Heuristic processing	Analytic processing
Evans & Over (1996)	Tacit thought processes	Explicit thought processes
Evans & Wason (1976); Wason & Evans (1975)	Type 1 processes	Type 2 processes
Fodor (1983)	Modular processes	Central processes
Gawronski & Bodenhausen (2006)	Associative processes	Propositional processes
Haidt (2001)	Intuitive system	Reasoning system
Johnson-Laird (1983)	Implicit inferences	Explicit inferences
Kahneman & Frederick (2002, 2005)	Intuition	Reasoning
Lieberman (2003)	Reflexive system	Reflective system
Loewenstein (1996)	Visceral factors	Tastes
Metcalfe & Mischel (1999)	Hot system	Cool system
Norman & Shallice (1986)	Contention scheduling	Supervisory attentional system
Pollock (1991)	Quick and inflexible modules	Intellection
Posner & Snyder (1975)	Automatic activation	Conscious processing
Reber (1993)	Implicit cognition	Explicit learning
Shiffrin & Schneider (1977)	Automatic processing	Controlled processing
Sloman (1996)	Associative system	Rule-based system
Smith & DeCoster (2000)	Associative processing	Rule-based processing
Strack & Deutsch (2004)	Impulsive system	Reflective system
Thaler & Shefrin (1981)	Doer	Planner
Toates (2006)	Stimulus-bound	Higher order
Wilson (2002)	Adaptive unconscious	Conscious

Evans, 1984, 1989). However, to attenuate the proliferation of nearly identical theories, I suggested the more generic terms System 1 and System 2 in a previous book (Stanovich, 1999). Although these terms have become popular, there is an infelicitousness to the System 1/System 2 terminology. Such terminology seems to connote that the two processes in dual-process theory map explicitly to two distinct brain systems. This is a stronger assumption than most theorists wish to make. Additionally, both Evans (2008, 2009, 2010; Chapter 8, this volume) and Stanovich (2004, 2011) have discussed how terms such as *System 1* or *heuristic system* are really misnomers because they imply that what is being referred to is a singular system. 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 higher level cognitive control. I have suggested (Stanovich, 2004) the acronym TASS (standing for The Autonomous Set of Systems) to describe what is in actuality a heterogeneous set.

Using the acronym TASS was a step forward in clearing up some of the confusion surrounding autonomous processes. For similar reasons, Evans (2008, 2009, Chapter 8; see also Samuels, 2009) has suggested a terminology of Type 1 processing versus Type 2 processing. The Type 1/Type 2 terminology captures better than previous terminology that a dual-*process* theory is not necessarily a dual-*system* theory (see Evans, 2008, 2009, for an extensive discussion). For these reasons, I will rely most heavily on the Type 1/Type 2 terminology. An even earlier terminology due to Evans (1984, 1989)—heuristic versus analytic processing—will also be employed on occasions when it is felicitous.

The defining feature of Type 1 processing is its autonomy—the execution of Type 1 processes is mandatory when their triggering stimuli are encountered, and they are not dependent on input from high-level control systems. Autonomous processes have other correlated features—their execution is rapid, they do not put a heavy load on central processing capacity, they tend to operate in parallel without interfering with themselves or with Type 2 processing—but these other correlated features are not defining. Autonomous processes would include behavioral regulation by the emotions; the encapsulated modules for solving specific adaptive problems that have been posited by evolutionary psychologists; processes of implicit learning; and the automatic firing of overlearned associations. Type 1 processes conjoin the properties of automaticity, quasi-modularity, and heuristic processing as these constructs have been variously discussed in cognitive science (Barrett & Kurzban, 2006; Carruthers, 2006; Coltheart, 1999; Evans, 2008, 2009; Moors & De Houwer, 2006; Samuels, 2005, 2009; Shiffrin & Schneider, 1977; Sperber, 1994).

It is important to emphasize that Type 1 processing is not limited to modular subprocesses that meet all of the classic Fodorian (1983) criteria. Type 1 processing encompasses processes of unconscious implicit learning and conditioning. Also, many rules, stimulus discriminations, and decision-making principles that have been practiced to automaticity (e.g., Kahneman & Klein, 2009; Shiffrin & Schneider, 1977) are processed in a Type 1 manner. This learned information can sometimes be just as much a threat to rational behavior 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 & Ayton, 1999; Hsee & Hastie, 2006).

Type 2 processing is nonautonomous. Type 2 processing contrasts with Type 1 processing on each of the correlated properties that define the latter. It is relatively slow and computationally expensive. Many Type 1 processes can operate at once in parallel, but Type 2 processing is largely serial. Type 2 processing is often language based, but it is not necessarily so. One of the most critical functions of Type 2 processing is to override Type 1 processing. This is sometimes necessary because autonomous processing has heuristic qualities. It is designed to get the response into the right ballpark when solving a problem or making a decision, but it is not designed for the type of fine-grained analysis called for in situations of unusual importance (financial decisions, fairness judgments, employment decisions, legal judgments, etc.). Heuristics depend on benign environments. In hostile environments, they can be costly (see Hilton, 2003; Over, 2000; Stanovich, 2004, 2009b). A benign environment means one that contains useful (that is, diagnostic) cues that can be exploited by various heuristics (for example, affect-triggering cues, vivid and salient stimulus components, convenient and accurate anchors). Additionally, for an environment to be classified as benign, it also must contain no other individuals who will adjust their behavior to exploit those relying only on heuristics. In contrast, a hostile environment for heuristics is one in which there are few cues that are usable by heuristic processes or there are misleading cues (Kahneman & Klein, 2009). Another way that an environment can turn hostile for a heuristic processor is if other agents discern the simple cues that are being used and the other agents start to arrange the cues for their own advantage (for example, advertisements, or the deliberate design of supermarket floor space to maximize revenue).

All of the different kinds of Type 1 processing (processes of emotional regulation, Darwinian modules, associative and implicit learning processes) can produce responses that are irrational in a particular context if not overridden. For example, often humans act as cognitive misers (see Stanovich, 2009b) by engaging in attribute substitution (Kahneman & Frederick, 2002)—the substitution of an easy-to-evaluate characteristic for a harder one, even if the easier one is less accurate. For example, the cognitive miser will substitute the less effortful attributes of vividness or affect for the more effortful retrieval of relevant facts (Kahneman, 2003; Li & Chapman, 2009; Slovic & Peters, 2006; Wang, 2009). But when we are evaluating important risks—such as the risk of certain activities and environments for our children—we do not want to substitute vividness for careful thought about the situation. In such situations, we want to employ Type 2 override processing to block the attribute substitution of the cognitive miser.

To override Type 1 processing, Type 2 processing must display at least two related capabilities. One is the capability of interrupting Type 1 processing and suppressing its response tendencies. Type 2 processing thus involves inhibitory mechanisms of the type that have been the focus of work on executive functioning (Aron, 2008; Best, Miller, & Jones, 2009; Hasher, Lustig, & Zacks, 2007; Miyake et al., 2000; Zelazo, 2004). But the ability to suppress Type 1 processing gets the job only half done. Suppressing one response is not helpful unless there is a better response available to substitute for it. Where do these better responses come from? One answer is that they come from processes of hypothetical reasoning and cognitive simulation that are a unique aspect of Type 2 processing (Johnson-Laird, Chapter 9). When we reason hypothetically, we create temporary models of the world and test out actions (or alternative causes) in that simulated world. To reason hypothetically we must, however, have one critical cognitive capability—we must be able to prevent our representations of the real world from becoming confused with representations of imaginary situations. The so-called cognitive decoupling operations are the central feature of Type 2 processing that make this possible, and they have implications for how we conceptualize both intelligence and rationality.

In a much-cited article, Leslie (1987) modeled pretense by positing a so-called secondary representation (see Perner, 1991) 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. The important issue for our purposes is that decoupling secondary representations from the world and then maintaining the decoupling while simulation is carried out is a Type 2 processing operation. It is computationally taxing and greatly restricts the ability to conduct any other Type 2 operation simultaneously. In fact, decoupling operations might well be a major contributor to a distinctive Type 2 property: its seriality.

Figure 22.1 represents a preliminary model of mind, based on what has been outlined thus far, with one important addition. The addition stems from the fact that instructions to initiate override of Type 1 processing (and to initiate simulation activities) must be controlled by cognitive machinery at a higher level than the decoupling machinery itself. Type 2 processing needs to be understood in terms of two levels of cognitive control—what are termed in Figure 22.1 the algorithmic level and the reflective level. There I have presented the tripartite proposal in the spirit of Dan Dennett's (1996) book *Kinds of Minds* where he used that title to suggest that within the brain of humans are control systems of very different types—different kinds of minds. I have labeled the traditional source of Type 1 processing as the autonomous mind but differentiated Type 2 processing into the algorithmic mind and the reflective mind. The autonomous mind can be overridden by algorithmic-level mechanisms; but override itself is initiated by higher level control. That is, the algorithmic level is conceptualized as subordinate to the higher level goal states and epistemic thinking dispositions of the reflective mind.

Individual Differences Within the Tripartite Model of Mind

Psychometricians have long distinguished typical performance situations from optimal (sometimes termed *maximal*) performance situations (see Ackerman, 1994, 1996; Ackerman & Heggestad,

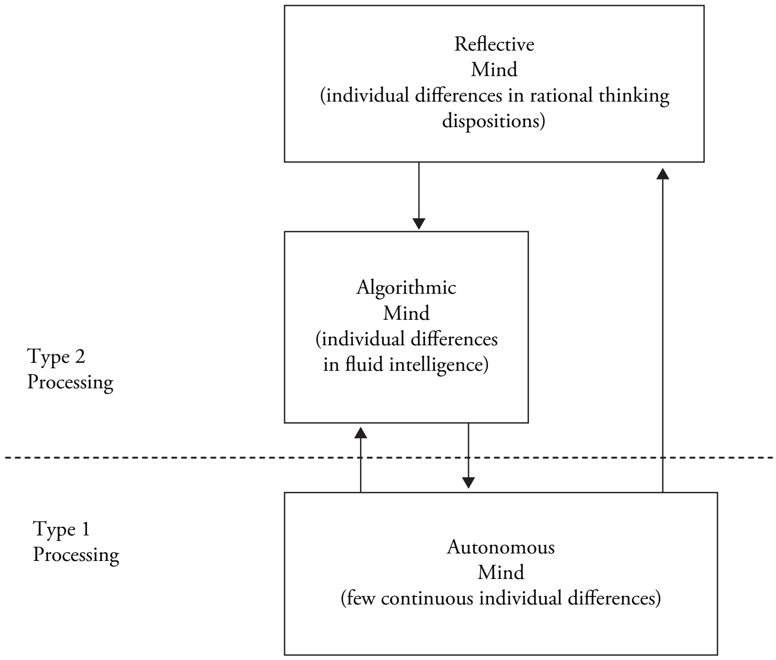

Fig. 22.1 The tripartite structure and the locus of individual differences.

1997; Ackerman & Kanfer, 2004; see also Cronbach, 1949; Matthews, Zeidner, & Roberts, 2002; Sternberg, Grigorenko, & Zhang, 2008). 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. The goals to be pursued in the task are left somewhat open. The issue is what a person would typically do in such a situation, given few constraints. Typical performance measures are measures of the reflective mind—they assess in part goal prioritization and epistemic regulation. In contrast, optimal performance situations are those where the task interpretation is determined externally. The person performing the task is instructed to maximize performance. Thus, optimal performance measures examine questions of the efficiency of goal pursuit—they capture the processing efficiency of the algorithmic mind. 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.

The difference between the algorithmic mind and the reflective mind is captured in another well-established distinction in the measurement of individual differences: the distinction between cognitive ability and thinking dispositions. The former are, as just mentioned, measures of the efficiency of the algorithmic mind. The latter travel under a variety of names in psychology—*thinking dispositions* or *cognitive styles* being the two most popular. Many thinking dispositions concern beliefs, belief structure and, importantly, attitudes toward forming and changing beliefs. Other thinking dispositions that have been identified concern a person's goals and goal hierarchy. Examples of some thinking dispositions that have been investigated by psychologists are actively open-minded thinking, need for cognition (the tendency to think a lot), consideration of future consequences, need for closure, superstitious thinking, and dogmatism (Cacioppo et al., 1996; Kruglanski & Webster, 1996; Norris & Ennis, 1989; Schommer-Aikins, 2004; Stanovich, 1999, 2009b; Sternberg, 2003; Sternberg & Grigorenko, 1997; Strathman et al., 1994).

The types of cognitive propensities that these thinking disposition measures reflect are the tendency to collect information before making up one's mind, the tendency to seek various points of view before coming to a conclusion, the disposition to think extensively about a problem before responding, the tendency to calibrate the degree of strength of one's opinion to the degree of evidence available, the tendency to think about future consequences before taking action, the tendency to explicitly weigh pluses and minuses of situations before making a decision, and the tendency to seek nuance and avoid absolutism. In short, individual differences in thinking dispositions are assessing variation in people's goal management, epistemic values, and epistemic self-regulation—differences in the operation of reflective mind (see Koedinger & Roll, Chapter 40). They are psychological characteristics that underpin rational thought and action.

The cognitive abilities assessed on intelligence tests are not of this type. They are not about high-level personal goals and their regulation, or about the tendency to change beliefs in the face of contrary evidence, or about how knowledge acquisition is internally regulated when not externally directed. People have indeed come up with *definitions* of intelligence that encompass such things. Theorists often define intelligence in ways that encompass rational action and belief but, nevertheless, *the actual measures of intelligence in use assess only algorithmic-level cognitive capacity*. No current intelligence test that is even moderately used in practice assesses rational thought or behavior (Stanovich, 2002, 2009b).

Figure 22.1 represents the classification of individual differences in the tripartite view. The broken horizontal line represents the location of the key distinction in older, dual-process views. Whereas the reflective and algorithmic minds are characterized by continuous individual differences and substantial variability, there are fewer continuous individual differences in the autonomous mind and less variability (see Kaufman et al., 2010, for a different view). Disruptions to the autonomous mind often reflect damage to cognitive modules that result in very discontinuous cognitive dysfunction such as autism or the agnosias and alexias (Anderson, 2005; Bermudez, 2001; Murphy & Stich, 2000).

Figure 22.1 identifies variation in fluid intelligence (Gf) with individual differences in the efficiency of processing of the algorithmic mind. Fluid intelligence is one component in the Cattell/Horn/Carroll (CHC) theory of intelligence (Carroll, 1993; Cattell, 1963, 1998; Horn & Cattell, 1967). Sometimes called the theory of fluid and crystallized intelligence (symbolized Gf/Gc theory), this theory posits that tests of mental ability tap, in addition to a general factor, a small number of broad factors, of which two are dominant (Geary, 2005; Horn & Noll, 1997; Taub & McGrew, 2004). Fluid intelligence (Gf) reflects reasoning abilities operating across a variety of domains—in particular, novel ones. It is measured by tasks of abstract reasoning such as figural analogies, Raven matrices, and series completion. Crystallized intelligence (Gc) reflects declarative knowledge acquired from acculturated learning experiences. It is measured by vocabulary tasks, verbal comprehension, and general knowledge measures. Ackerman (1996) discusses how the two dominant factors in the CHC theory reflect a long history of considering two aspects of intelligence: intelligence-as-process (Gf) and intelligence-as-knowledge (Gc).

I have argued that individual differences in fluid intelligence are a key indicator of the variability across individuals in the ability to sustain decoupling operations (Stanovich, 2001, 2009b). Increasingly it is becoming apparent that one of the critical mental operations being tapped by measures of fluid intelligence is the cognitive decoupling operation I have discussed in this chapter. This is becoming clear from converging work on executive function and working memory. Most measures of executive function and working memory are direct or indirect indicators of a person's ability to sustain decoupling operations (Duncan et al., 2008; Engle, 2002; Gray, Chabris, & Braver, 2003; Hasher, Lustig, & Zacks, 2007; Kane, 2003; Lepine, Barrouillet, & Camos, 2005; Salthouse, Atkinson, & Berish, 2003; Salthouse & Pink, 2008; Stanovich, 2011).

Figure 22.1 highlights an important sense in which rationality is a more encompassing construct than intelligence. As previously discussed, to be rational, a person must have well-calibrated beliefs and must act appropriately on those beliefs to achieve goals—both properties of the reflective mind. The person must, of course, have the algorithmic-level machinery that enables him or her 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 fluid intelligence (the algorithmic mind) or because of individual differences in thinking dispositions (the reflective mind).

The conceptualization in Figure 22.1 has several advantages. First, it conceptualizes intelligence in terms of what intelligence tests actually measure. IQ tests do not attempt to measure directly an aspect of epistemic or instrumental rationality, nor do they examine any thinking dispositions that relate to rationality. It is also clear from Figure 22.1 why rationality and intelligence can become dissociated. Rational thinking depends on thinking dispositions as well as algorithmic efficiency. Thus, as long as variation in thinking dispositions is not perfectly correlated with fluid intelligence, there is the statistical possibility of dissociations between rationality and intelligence.

In fact, substantial empirical evidence indicates that individual differences in thinking dispositions and intelligence are far from perfectly correlated. Many different studies involving thousands of subjects (e.g., Ackerman & Heggestad, 1997; Austin & Deary, 2002; Baron, 1982; Bates & Shieles, 2003; Cacioppo et al., 1996; Eysenck, 1994; Goff & Ackerman, 1992; Kanazawa, 2004; Kokis et al., 2002; Zeidner & Matthews, 2000) have indicated that measures of intelligence display only moderate to weak correlations (usually less than .30) with some thinking dispositions (e.g., actively open-minded thinking, need for cognition) and near-zero correlations with others (e.g., conscientiousness, curiosity, diligence). Other important evidence supports the conceptual distinction made here between algorithmic cognitive capacity and thinking dispositions. For example, across a variety of tasks from the heuristics and biases literature, it has consistently been found that rational thinking dispositions will predict variance after the effects of general intelligence have been controlled.[4]

The functions of the different levels of control are illustrated more completely in Figure 22.2. There, it is clear that the override capacity itself is a property of the algorithmic mind and it is indicated by the arrow labeled A. However, previous dual-process theories have tended to ignore the higher level cognitive function that initiates the override function in the first place. This is a dispositional property of the reflective mind that is related to rationality. In the model in Figure 22.2, it corresponds to arrow B, which represents (in machine intelligence terms) the call to the algorithmic mind to override the Type 1 response by taking it offline. This is a different mental function than the override function itself (arrow A), and there, the evidence cited earlier indicates that the two functions are indexed by different types of individual differences.

Figure 22.2 represents another aspect of cognition somewhat neglected by previous dual-process theories. Specifically, 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 22.2 explicitly represents the simulation function as well as the fact that the call to initiate simulation originates in the reflective mind. The decoupling

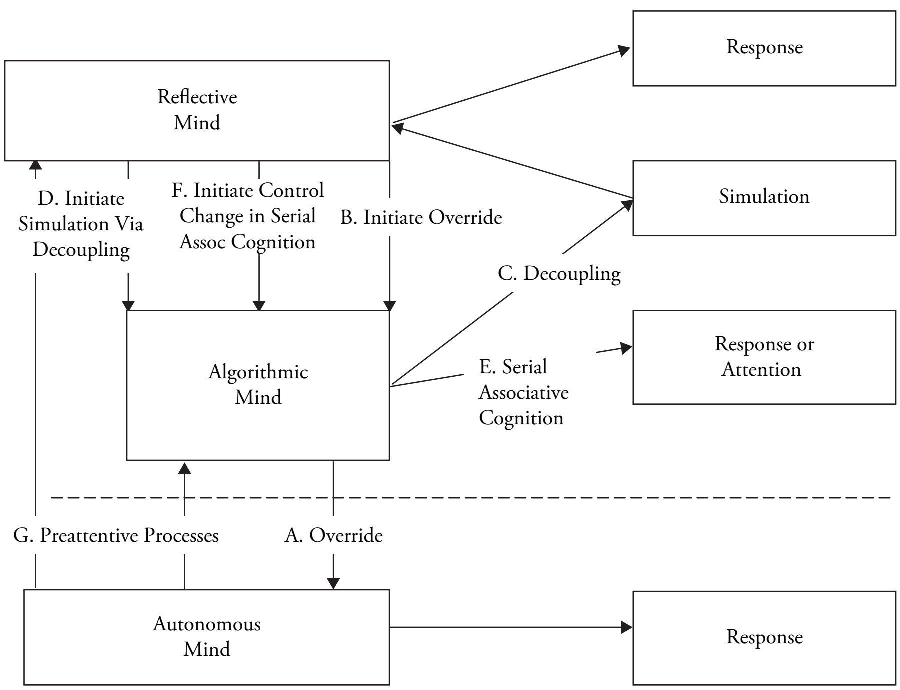

Fig. 22.2 A more complete model of the tripartite structure.

operation (indicated by arrow C) itself is carried out by the algorithmic mind and the call to initiate simulation (indicated by arrow D) by the reflective mind. Again, 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. Also represented is the fact that the higher levels of control receive inputs from the computations of the autonomous mind (arrow G) via so-called preattentive processes (Evans, 2006, 2007, 2008, 2009). The arrows labeled E and F reflect the decoupling and higher level control of a kind of Type 2 processing (serial associative cognition) that does not involve fully explicit cognitive simulation (see Stanovich, 2011).

Mindware in the Tripartite Model

Knowledge bases, both innate and derived from experience, also importantly bear on rationality. The term *mindware* was coined by Perkins (1995) to refer to the rules, knowledge, procedures, and strategies that a person can retrieve from memory in order to aid decision making and problem solving. Each of the levels in the tripartite model of mind has to access knowledge to carry out its operations, as illustrated in Figure 22.3. As the Figure indicates, the reflective mind not only accesses general knowledge structures (Gc) but, importantly, accesses the person's opinions, beliefs, and reflectively acquired goal structure. The algorithmic mind accesses microstrategies for cognitive operations and production system rules for sequencing behaviors and thoughts. Finally, the autonomous mind accesses not only knowledge bases that are evolutionary adaptations, but it also retrieves information that has become tightly compiled and available to the autonomous mind due to overlearning and practice.

It is important to note that what is displayed in Figure 22.3 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 (see arrow G in Figure 22.2). The mindware available for retrieval, particularly that available to the reflective mind, is in part the product of past learning experiences. The knowledge structures available for retrieval by the reflective mind represent Gc, crystallized intelligence. Recall that Gf, fluid intelligence (intelligence-as-process), is already represented in the Figure 22.2. It is the general computational power of the algorithmic

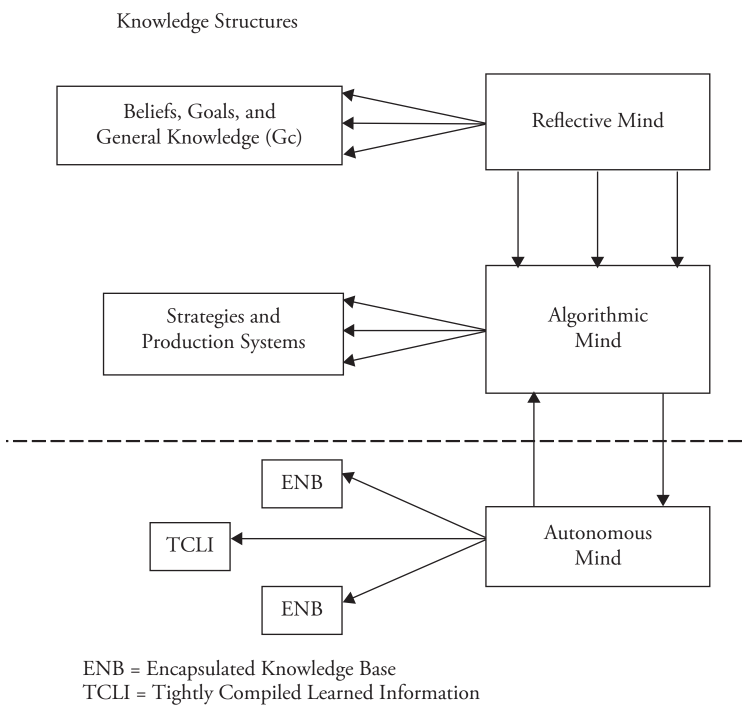

Fig. 22.3 Knowledge structures in the tripartite framework.

mind—importantly exemplified by the ability to sustain cognitive decoupling.

It is important to see how both of the major components of Gf/Gc theory miss critical aspects of rational thought. Fluid intelligence will, of course, have some relation to rationality because it indexes the computational power of the algorithmic mind to sustain decoupling. Because override and simulation are important operations for rational thought, Gf will definitely facilitate rational action in some situations. Nevertheless, the tendency to initiate override (arrow B in Fig. 22.2) and to initiate simulation activities (arrow D in Fig. 22.2) are both aspects of the reflective mind unassessed by intelligence tests, so the tests will miss these components of rationality.

The situation with respect to Gc is a little different. It is true that much of the mindware of rational thought would be classified as crystallized intelligence in the abstract. But is it the kind of crystallized knowledge that is specifically assessed on the tests? The answer is no. The mindware of rational thought is somewhat specialized mindware (see Stanovich, 2009b). It clusters in the domains of probabilistic reasoning (see Griffiths et al., Chapter 3), causal reasoning (see Cheng & Buehner, Chapter 12), and scientific reasoning (see Dunbar & Klahr, Chapter 35). In contrast, the crystallized knowledge assessed on IQ tests is deliberately designed to be nonspecialized. The designers of the tests, in order to make sure the sampling of Gc is fair and unbiased, explicitly attempt to *broadly* sample vocabulary, verbal comprehension domains, and general knowledge. The broad sampling ensures unbiasedness in the test, but it inevitably means that the specific knowledge bases critical to rationality will go unassessed. In short, Gc, as traditionally measured, does not assess individual differences in rationality, and Gf will do so only indirectly and to a mild extent. In short, as measures of rational thinking, IQ tests are radically incomplete.

The Requirements of Rational Thinking and Their Relation to Intelligence

Within the tripartite framework, rationality requires mental characteristics of three different types. Problems in rational thinking arise when cognitive capacity is insufficient to sustain autonomous system override, when the necessity of override is not recognized, or when simulation processes do not have access to the mindware necessary for the synthesis of a better response. The source of these problems, and their relation to intelligence, help to explain one data trend that has been uncovered—that some rational thinking problems show surprising degrees of dissociation from cognitive ability (Stanovich, 2009b, 2011; Stanovich & West, 2007, 2008a, 2008b; West, Toplak, & Stanovich, 2008). Myside bias, for example, is virtually independent of intelligence (Macpherson & Stanovich, 2007; Sá, Kelley, Ho, & Stanovich, 2005; Stanovich & West, 2007, 2008a, 2008b; Toplak & Stanovich, 2003). Individuals with higher IQs in a university sample are no less likely to process information from an egocentric perspective than are individuals with relatively lower IQs.

Irrational behavior can occur because the right mindware (cognitive rules, strategies, knowledge, and belief systems) is not available to use in decision making. We would expect to see a correlation with intelligence here because mindware gaps most often arise from lack of education or experience. Nevertheless, while it is true that more intelligent individuals learn more things than less intelligent individuals, much knowledge (and many thinking dispositions) relevant to rationality are picked up rather late in life. Explicit teaching of this mindware is not uniform in the school curriculum at any level. That such principles are taught very inconsistently means that some intelligent people may fail to learn these important aspects of critical thinking. In university samples, correlations with cognitive ability have been found to be roughly (in absolute magnitude) in the range of .20–.35 for probabilistic reasoning tasks and scientific reasoning tasks measuring a variety of rational principles (Bruine de Bruin, Parker, & Fischhoff, 2007; Kokis et al., 2002; Parker & Fischhoff, 2005; Sá, West, & Stanovich, 1999; Stanovich & West, 1998b, 1998c, 1998d, 1999, 2000, 2008b; Toplak & Stanovich, 2002). This is again a magnitude of correlation that allows for substantial discrepancies between intelligence and rationality. Intelligence is thus no inoculation against many of the sources of irrational thought. None of these sources are directly assessed on intelligence tests, and the processes that *are* tapped by IQ tests are not highly overlapping with the processes and knowledge that explain variation in rational thinking ability.

Conclusions and Future Directions

The many studies of individual differences on heuristics and biases tasks falsify the most explicit versions of the Panglossian view. People are not all identically rational. There are individual differences on all of these tasks and the individual differences are not the result of random performance errors (see Stanovich, 1999; Stein, 1996). Instead, they are systematic. If there are such systematic individual differences, it means that at least some people, some of the time, are irrational. Extreme Panglossianism cannot be sustained. However, there is another position in the debate that, like Panglossianism, serves to minimize the attribution of irrationality. Termed the Apologist position (Stanovich, 1999), this view takes very seriously the idea that humans have computational limitations that keep them from being fully rational.

Like the Meliorist, the Apologist accepts the empirical reality and nonspuriousness of normative/descriptive gaps, but the Apologist is much more hesitant to term them instances of irrationality. This is because the Apologist takes the position that to characterize a suboptimal behavior as irrational, it must be the case that the normative model is computable by the individual. If there are computational limitations affecting task performance, then the normative model may not be prescriptive, at least for individuals of low algorithmic capacity. Prescriptive models are usually viewed as specifying how processes of belief formation and decision making should be carried out, given the limitations of the human cognitive apparatus and the situational constraints (e.g., time pressure) with which the decision maker must deal (Baron, 2008). From the Apologist's perspective, the descriptive model is quite close to the prescriptive model and the descriptive/normative gap is attributed to a computational limitation. Although the Apologist admits that performance is suboptimal from the standpoint of the normative model, it is not irrational because there is no prescriptive/descriptive gap.

However, as demonstrated in some of the data patterns I have described in this chapter, the Apologist stratagem will not work for all of the irrational tendencies that have been uncovered in the heuristics and biases literature. This is because many biases are not very strongly correlated with measures of intelligence (algorithmic capacity). Additionally, there is reliable variance in rational thinking found even after cognitive ability is controlled, and that reliable variance is associated with thinking dispositions in theoretically predictable ways. These thinking dispositions reflect control features of the reflective mind that can lead to responses that are more or less rational. They are one of the main sources of the individual differences in rational thought that I have been exploring in this chapter. Such thinking dispositions vary systematically from individual to individual, and they are the source of what Meliorists consider the variance in the irrationalities in human cognition. Unlike the Panglossian (who assumes uniform rationality) or the Apologist (who minimizes such variability while not entirely denying it), the Meliorist is very accepting of the idea of variability in rational thought.

The Panglossian position in the Great Rationality Debate has obscured the existence of individual differences in rational thought and its underlying components. In particular, Panglossian philosophers have obscured the importance of the reflective mind. Philosophical treatments of rationality by Panglossians tend to have a common structure (see Cohen, 1981, for example). Such treatments tend to stress the importance of the competence/performance distinction and then proceed to allocate all of the truly important psychological mechanisms to the competence side of the dichotomy.

For example, Rescher (1988) argues that "to construe the data of these interesting experimental studies [of probabilistic reasoning] to mean that people are systematically programmed to fallacious processes of reasoning—rather than merely that they are inclined to a variety of (occasionally questionable) substantive suppositions—is a very questionable step . . . . While all (normal) people are to be credited with the capacity to reason, they frequently do not exercise it well" (p. 196). There are two parts to Rescher's (1988) point here: the "systematically programmed" part and the "inclination toward questionable suppositions" part (or, as Rips (1994, p. 394) puts it, whether incorrect reasoning is "systematically programmed or just a peccadillo"). Rescher's (1988) focus—like that of many who have dealt with the philosophical implications of the idea of human irrationality—is on the issue of how humans are "systematically programmed." "Inclinations toward questionable suppositions" are only of interest to those in the philosophical debates as mechanisms that allow one to drive a wedge between competence and performance—thus maintaining a theory of near-optimal human rational competence in the face of a host of responses that seemingly defy explanation in terms of standard normative models.

Analogously to Rescher, Cohen (1982) argues that there really are only two factors affecting performance on rational thinking tasks: "normatively correct mechanisms on the one side, and adventitious causes of error on the other" (p. 252). Not surprisingly given such a conceptualization, the processes contributing to error ("adventitious causes") are of little interest to Cohen (1981, 1982). In his view, human performance arises from an intrinsic human competence that is impeccably rational, but responses occasionally deviate from normative correctness due to inattention, memory lapses, lack of motivation, and other fluctuating but basically unimportant causes (in Cohen's view). There is nothing in such a conception that would motivate any interest in patterns of errors or individual differences in such errors.

One of the goals of this chapter is to reverse the figure and ground in the rationality debate, which has tended to be dominated by the particular way that philosophers frame the competence/performance distinction. From a psychological standpoint, there may be important implications in precisely the aspects of performance that have been backgrounded in this controversy ("adventitious causes," "peccadillos"). That is, whatever the outcome of the disputes about how humans are "systematically programmed," variation in the "inclination toward questionable suppositions" is of psychological interest as a topic of study in its own right. The research discussed in this chapter provides at least tentative indications that the "inclination toward questionable suppositions" has some degree of domain generality and that it is predicted by thinking dispositions that concern the epistemic and pragmatic goals of the individual and that are part of the reflective mind.

Johnson-Laird and Byrne (1993; see Johnson-Laird, 2006) articulate a view of rational thought that parses the competence/performance distinction much differently from that of Cohen (1981, 1982, 1986). It is a view that highlights the importance of the reflective mind and leaves room for individual differences in important components of cognition. At the heart of the rational competence that Johnson-Laird and Byrne (1993) attribute to humans is not perfect rationality but instead just one meta-principle: People are programmed to accept inferences as valid provided that they have constructed no mental model of the premises that contradict the inference (see Johnson-Laird, Chapter 9). Inferences are categorized as false when a mental model is discovered that is contradictory. However, the search for contradictory models is "not governed by any systematic or comprehensive principles" (p. 178).

The key point in Johnson-Laird and Byrne's (1993) account is that once an individual constructs a mental model from the premises, once the individual draws a new conclusion from the model, and once the individual begins the search for an alternative model of the premises that contradicts the conclusion, the individual "lacks any systematic method to make this search for counter-examples" (p. 205). Here is where Johnson-Laird and Byrne's (1993) model could be modified to allow for the influence of thinking styles in ways that the impeccable competence view of Cohen (1981) does not. In this passage, Johnson-Laird and Byrne seem to be arguing that there are no systematic control features of the search process. But epistemically related thinking dispositions may in fact be reflecting just such control features.

Individual differences in the extensiveness of the search for contradictory models could arise from a variety of cognitive factors that, although they may not be completely systematic, may be far from "adventitious"—factors such as dispositions toward premature closure, cognitive confidence, reflectivity, dispositions toward confirmation bias, and ideational generativity. The decontextualizing requirement of many heuristics and biases tasks is a feature that is emphasized by many critics of that literature who, nevertheless, fail to see it as implying a research program for differential psychology. For example, I have argued that to contextualize a problem is such a ubiquitous reasoning style for human beings that it constitutes one of a very few so-called fundamental computational biases of information processing (Stanovich, 2003, 2004). Thus, it is not surprising that many people respond incorrectly when attempting a psychological task that is explicitly designed to require a decontextualized reasoning style (contrary-to-fact syllogisms, argument evaluation, etc.). But recall the empirically demonstrated variability on all of these tasks. The fact that some people *do* give the decontextualized response means that at least some people have available a larger repertoire of reasoning styles, allowing them to reason flexibly reason so as to override fundamental computational biases if the situation requires.

Another way of stressing the importance of individual differences in understanding the nature of rational thought is in terms of Dennett's (1987) so-called intentional stance, which he marries to an assumption of idealized rationality. Dennett (1988) argues that we use the intentional stance for humans and dogs but not for lecterns because for the latter "there is no predictive leverage gained by adopting the intentional stance" (p. 496). However, in several experiments discussed in this chapter, it has been shown that there is additional predictive leverage to be gained by relaxing the idealized rationality assumption of Dennett's (1987, 1988) intentional stance and by positing measurable and systematic variation in intentional-level psychologies (that is, in the reflective mind). Knowledge about such individual differences in people's intentional-level psychologies can be used to predict variance in the normative/descriptive gap displayed on many reasoning tasks. Consistent with the Meliorist conclusion that there can be individual differences in human rationality, the results show that there is variability in reasoning that cannot be accommodated within a model of perfect rational competence operating in the presence of performance errors and computational limitations.

Acknowledgments

This research was supported by grants from the Social Sciences and Humanities Research Council of Canada and the Canada Research Chairs program to Keith E. Stanovich. Laura Noveck and Keith Holyoak are thanked for their insightful comments on an earlier version of this chapter.

Notes

1. It should also be noted that in the view of rationality taken in this chapter, rationality is an intentional-level personal entity and not an algorithmic-level subpersonal one (Bermudez, 2001; Davies, 2000; Frankish, 2009; Stanovich, 1999, 2009a). A memory system in the human brain is not rational or irrational—it is merely efficient or inefficient (or of high or low capacity). Thus, subprocesses of the brain do not display rational or irrational properties per se, although they may contribute in one way or another to personal decisions or beliefs that could be characterized as such. Rationality concerns the actions of an entity in its environment that serve its goals. One of course could extrapolate the notion of environment to include the interior of the brain itself and then talk of a submodule that chose strategies rationally or not. This move creates two problems. First, what are the goals of this subpersonal entity—what are its interests that its rationality is trying to serve? This is unclear in the case of a subpersonal entity. Second, such a move regresses all the way down. We would need to talk of a neuron firing being either rational or irrational. As Oaksford and Chater (1998) put it, "the fact that a model is optimizing something does not mean that the model is a rational model. Optimality is not the same as rationality….Stomachs may be well or poorly adapted to their function (digestion), but they have no beliefs, desires or knowledge, and hence the question of their rationality does not arise" (pp. 4 and 5).

2. The principle of maximizing expected value says that the action that a rational person should choose is the one with the highest expected value. Expected value is calculated by taking the objective value of each outcome and multiplying it by the probability of that outcome and then summing those products over all of the possible outcomes. Symbolically, the formula is as follows: Expected value = $\sum p_i v_i$; where p_i is the probability of each outcome and v_i is the value of each outcome. The symbol $\sum$ is the summation sign, and simply means "add up all of the terms that follow." The term *utility* refers to subjective value. Thus, the calculation of expected utility involves identical mathematics except that a subjective estimate of utility is substituted for the measure of objective value.

3. It is important to note that the Meliorist recognizes two different ways in which human decision-making performance might be improved. These might be termed *cognitive change* and *environmental change*. First, it might be possible to teach people better reasoning strategies and to have them learn rules of decision making that are helpful (see Stanovich, 2009b). These would represent instances of cognitive change. Additionally, however, research has shown that it is possible to change the environment so that natural human reasoning strategies will not lead to error (Gigerenzer, 2002; Milkman, Chugh, & Bazerman, 2009; Thaler & Sunstein, 2008). For example, choosing the right default values for a decision would be an example of an environmental change. In short, environmental alterations (as well as cognitive changes) can prevent rational thinking problems. Thus, in cases where teaching people the correct reasoning strategies might be difficult, it may well be easier to change the environment so that decision-making errors are less likely to occur.

4. Such empirical studies indicate that cognitive capacity and thinking dispositions measures are tapping separable variance. The converging evidence on the existence of this separable variance is growing (Bruine de Bruin, Parker, & Fischhoff, 2007; Finucane & Gullion, 2010; Klaczynski, Gordon, & Fauth, 1997; Klaczynski & Lavallee, 2005; Klaczynski & Robinson, 2000; Kokis et al., 2002; Macpherson & Stanovich, 2007; Newstead, Handley, Harley, Wright, & Farrelly, 2004; Parker & Fischhoff, 2005; Sá & Stanovich, 2001; Stanovich & West, 1997, 1998c, 2000; Toplak, Liu, Macpherson, Toneatto, & Stanovich, 2007; Toplak & Stanovich, 2002).

References

Ackerman, P. L. (1994). Intelligence, attention, and learning: Maximal and typical performance. In D. K. Detterman (Ed.), *Current topics in human intelligence* (Vol. 4, pp. 1–27). Norwood, NJ: Ablex.

Ackerman, P. L. (1996). A theory of adult development: Process, personality, interests, and knowledge. *Intelligence*, *22*, 227–257.

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*, 219–245.

Ackerman, P. L., & Kanfer, R. (2004). Cognitive, affective, and conative aspects of adult intellect within a typical and maximal performance framework. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (119–141). Mahwah, NJ: Lawrence Erlbaum Associates.

Ainslie, G. (1982). A behavioral economic approach to the defence mechanisms: Freud's energy theory revisited. *Social Science Information*, *21*, 735–780.

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

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, M. (2005). Marrying intelligence and cognition: A developmental view. In R. J. Sternberg & J. E. Pretz

(Eds.), *Cognition and intelligence* (pp. 268–287). New York: Cambridge University Press.

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

Aron, A. R. (2008). Progress in executive-function research: From tasks to functions to regions to networks. *Current Directions in Psychological Science*, *17*, 124–129.

Audi, R. (1993). *The structure of justification*. Cambridge, England: Cambridge University Press.

Audi, R. (2001). *The architecture of reason: The structure and substance of rationality*. Oxford, England: Oxford University Press.

Aunger, R., & Curtis, V. (2008). Kinds of behaviour. *Biology and Philosophy*, *23*, 317–345.

Austin, E. J., & Deary, I. J. (2002). Personality dispositions. In R. J. Sternberg (Ed.), *Why smart people can be so stupid* (pp. 187–211). New Haven, CT: Yale University Press.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*, 462–479.

Baron, J. (1982). Personality and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 308–351). Cambridge, England: Cambridge University Press.

Baron, J. (2008). *Thinking and deciding* (4th ed.). New York: Cambridge University Press.

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review*, *113*, 628–647.

Bates, T. C., & Shieles, A. (2003). Crystallized intelligence as a product of speed and drive for experience: the relationship of inspection time and openness to g and Gc. *Intelligence*, *31*, 275–287.

Bazerman, M., & Moore, D. A. (2008). *Judgment in managerial decision making*. New York: John Wiley.

Bazerman, M., Tenbrunsel, A., & Wade-Benzoni, K. (1998). Negotiating with yourself and losing: Understanding and managing conflicting internal preferences. *Academy of Management Review*, *23*, 225–241.

Bermudez, J. L. (2001). Normativity and rationality in delusional psychiatric disorders. *Mind and Language*, *16*, 457–493.

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review*, *29*, 180–200.

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

Bishop, M. A., & Trout, J. D. (2005). *Epistemology and the psychology of human judgment*. Oxford, England: Oxford University Press.

Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience. In H. W. Reese & R. Kail (Eds.), *Advances in child development and behavior* (Vol. 28, pp. 41–100). San Diego, CA: Academic Press.

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology*, *92*, 938–956.

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

Camerer, C., Loewenstein, G., & Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, *34*, 9–64.

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

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

Cattell, R. B. (1963). Theory for fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*, 1–22.

Cattell, R. B. (1998). Where is intelligence? Some answers from the triadic theory. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (29–38). Mahwah, NJ: Erlbaum.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). New York: Guilford Press.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, *4*, 317–370.

Cohen, L. J. (1982). Are people programmed to commit fallacies? Further thoughts about the interpretation of experimental data on probability judgment. *Journal for the Theory of Social Behavior*, *12*, 251–274.

Cohen, L. J. (1986). *The dialogue of reason*. Oxford, England: Oxford University Press.

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, *4*, 20–33.

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences*, *3*, 115–120.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*, 1–73.

Cronbach, L. J. (1949). *Essentials of psychological testing*. New York: Harper.

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

Dawes, R. M. (1998). Behavioral decision making and judgment. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, pp. 497–548). Boston, MA: McGraw-Hill.

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

Del Missier, F., Mantyla, T., & Bruine de Bruin, W. (2010). Executive functions in decision making: An individual differences approach. *Thinking and Reasoning*, *16*, 69–97.

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

Dennett, D. C. (1988). Precis of "The Intentional Stance". *Behavioral and Brain Sciences*, *11*, 493–544.

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.

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

Doherty, M. (2003). Optimists, pessimists, and realists. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 643–679). New York: Cambridge University Press.

Dohmen, T., Falk, A., Huffman, D., Marklein, F., & Sunde, U. (2009). Biased probability judgment: Evidence of incidence and relationship to economic outcomes from a representative sample. *Journal of Economic Behavior and Organization*, *72*, 903–915.
Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., . . . Nimmo-Smith, I. (2008). Goal neglect and Spearman's g: Competing parts of a complex task. *Journal of Experimental Psychology-General*, *137*, 131–148.
Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*, 380–417.
Edwards, W., & von Winterfeldt, D. (1986). On cognitive illusions and their implications. In H. R. Arkes & K. R. Hammond (Eds.), *Judgment and decision making* (pp. 642–679). Cambridge, England: Cambridge University Press.
Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19–23.
Evans, J. St. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, *75*, 451–468.
Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Hove, England: Erlbaum.
Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*, 454–459.
Evans, J. St. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin and Review*, *13*, 378–395.
Evans, J. St. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgment*. New York: Psychology Press.
Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology*, *59*, 255–278.
Evans, J. St. B. T. (2009). How many dual-process theories do we need? One, two, or many? In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 33–54). Oxford, England: Oxford University Press.
Evans, J. St. B. T. (2010). *Thinking twice: Two minds in one brain*. Oxford, England: Oxford University Press.
Evans, J. St. B. T., & Frankish, K. (Eds.). (2009). *In two minds: Dual processes and beyond*. Oxford, England: Oxford University Press.
Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove, England: Erlbaum.
Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, England: Psychology Press.
Evans, J. St. B. T., & Over, D. E. (2010). Heuristic thinking and human intelligence: A commentary on Marewski, Gaissmaier and Gigerenzer. *Cognitive Processing*, *11*, 171–175.
Evans, J. St. B. T., & Wason, P. C. (1976). Rationalization in a reasoning task. *British Journal of Psychology*, *67*, 479–486.
Eysenck, H. J. (1994). Personality and intelligence: Psychometric and experimental approaches. In R. J. Sternberg & P. Ruzgis (Eds.), *Personality and intelligence* (pp. 3–31). Cambridge, England: Cambridge University Press.
Finucane, M. L., & Gullion, C. M. (2010). Developing a tool for measuring the decision-making competence of older adults. *Psychology and Aging*, *25*, 271–288.
Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT University Press.
Foley, R. (1987). *The theory of epistemic rationality*. Cambridge, MA: Harvard University Press.
Frank, M. J., Cohen, M., & Sanfey, A. G. (2009). Multiple systems in decision making. *Current Directions in Psychological Science*, *18*, 73–77.
Frankish, K. (2004). *Mind and supermind*. Cambridge, England: Cambridge University Press.
Frankish, K. (2009). Systems and levels: Dual-system theories and the personal-subpersonal distinction. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 89–107). Oxford, England: Oxford University Press.
Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25–42.
Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731.
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*, 592–596.
Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster.
Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. New York: Viking Penguin.
Gilbert, D. T. (2006). *Stumbling on happiness*. New York: Alfred A. Knopf.
Gilboa, I. (2010). *Rational choice*. Cambridge, MA: The MIT Press.
Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.
Goff, M., & Ackerman, P. L. (1992). Personality-intelligence relations: Assessment of typical intellectual engagement. *Journal of Educational Psychology*, *84*, 537–552.
Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*, 316–322.
Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834.
Harman, G. (1995). Rationality. In E. E. Smith & D. N. Osherson (Eds.), *Thinking* (Vol. 3, pp. 175–211). Cambridge, MA: The MIT Press.
Hasher, L., Lustig, C., & Zacks, R. (2007). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (pp. 227–249). New York: Oxford University Press.
Hastie, R., & Dawes, R. M. (2010). *Rational choice in an uncertain world* (2nd ed.). Thousand Oaks, CA: Sage.
Hilton, D. J. (2003). Psychology and the financial markets: Applications to understanding and remedying irrational decision-making. In I. Brocas & J. D. Carrillo (Eds.), *The psychology of economic decisions. Vol. 1: Rationality and well-being* (pp. 273–297). Oxford, England: Oxford University Press.
Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, *26*, 1–23.
Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. Flanagan, J. Genshaft, & P. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). New York: Guilford Press.

Hsee, C. K., & Hastie, R. (2006). Decision and experience: Why don't we choose what makes us happy? *Trends in Cognitive Sciences*, *10*, 31–37.

Hurley, S., & Nudds, M. (2006). The questions of animal rationality: Theory and evidence. In S. Hurley & M. Nudds (Eds.), *Rational animals?* (pp. 1–83). Oxford, England: Oxford University Press.

Jeffrey, R. C. (1983). *The logic of decision (Second Ed.)*. Chicago, IL: University of Chicago Press.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N. (2006). *How we reason*. Oxford, England: Oxford University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (1993). Models and deductive rationality. In K. Manktelow & D. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 177–210). London: Routledge.

Jungermann, H. (1986). The two camps on rationality. In H. R. Arkes & K. R. Hammond (Eds.), *Judgment and decision making* (pp. 627–641). Cambridge, England: Cambridge University Press.

Kahneman, D. (2000). A psychological point of view: Violations of rational rules as a diagnostic of mental processes. *Behavioral and Brain Sciences*, *23*, 681–683.

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

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

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

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise a failure to disagree. *American Psychologist, 64*, 515–526.

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. (2006). Would you be happier if you were richer? A focusing illusion. *Science*, *312*, 1908–1910.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.

Kahneman, D., & Tversky, A. (1983). Can irrationality be intelligently discussed? *Behavioral and Brain Sciences*, *6*, 509–510.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*, 582–591.

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

Kanazawa, S. (2004). General intelligence as a domain-specific adaptation. *Psychological Review*, *111*, 512–523.

Kane, M. J. (2003). The intelligent brain in conflict. *Trends in Cognitive Sciences*, *7*, 375–377.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jimenez, L., Brown, J., & Mackintosh, N. J. (2010). Implicit learning as an ability. *Cognition*, *116*, 321–340.

Klaczynski, P. A. (2001). Analytic and heuristic processing influences on adolescent reasoning and decision making. *Child Development*, *72*, 844–861.

Klaczynski, P. A., Gordon, D. H., & Fauth, J. (1997). Goal-oriented critical reasoning and individual differences in critical reasoning biases. *Journal of Educational Psychology*, *89*, 470–485.

Klaczynski, P. A., & 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*, *92*, 1–24.

Klaczynski, P. A., & Robinson, B. (2000). Personal theories, intellectual ability, and epistemological beliefs: Adult age differences in everyday reasoning tasks. *Psychology and Aging*, *15*, 400–416.

Koehler, D. J., & James, G. (2009). Probability matching in choice under uncertainty: Intuition versus deliberation. *Cognition*, *113*, 123–127.

Koehler, D. J., & James, G. (2010). Probability matching and strategy availability. *Memory and Cognition*, *38*, 667–676.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative and methodological challenges. *Behavioral and Brain Sciences*, *19*, 1–53.

Kokis, J., Macpherson, R., Toplak, M., West, R. F., & 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., & 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*, 313–376.

Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing the mind: "Seizing" and "freezing". *Psychological Review*, *103*, 263–283.

Kuhberger, A. (2002). The rationality of risky decisions: A changing message. *Theory and Psychology*, *12*, 427–452.

Larrick, R. P. (2004). Debiasing. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 316–337). Malden, MA: Blackwell.

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

Lepine, R., Barrouillet, P., & Camos, V. (2005). What makes working memory spans so predictive of high-level cognition? *Psychonomic Bulletin and Review*, *12*, 165–170.

Leslie, A. M. (1987). Pretense and representation: The origins of "Theory of Mind". *Psychological Review*, *94*, 412–426.

Li, M., & Chapman, G. B. (2009). "100% of anything looks good:" The appeal of one hundred percent. *Psychonomic Bulletin and Review*, *16*, 156–162.

Lieberman, M. D. (2003). Reflexive and reflective judgment processes: A social cognitive neuroscience approach. In J. P. Forgas, K. R. Williams, & W. von Hippel (Eds.), *Social judgments: Implicit and explicit processes* (pp. 44–67). New York: Cambridge University Press.

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology*, *58*, 259–289.

Lieberman, M. D. (2009). What zombies can't do: A social cognitive neuroscience approach to the irreducibility of reflective consciousness. In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 293–316). Oxford, England: Oxford University Press.

Loewenstein, G. F. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, *65*, 272–292.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions*. New York: Wiley.
Macpherson, R., & Stanovich, K. E. (2007). Cognitive ability, thinking dispositions, and instructional set as predictors of critical thinking. *Learning and Individual Differences*, *17*, 115–127.
Manktelow, K. I. (2004). Reasoning and rationality: The pure and the practical. In K. I. Manktelow & M. C. Chung (Eds.), *Psychology of reasoning: Theoretical and historical perspectives* (pp. 157–177). Hove, England: Psychology Press.
Marewski, J. N., Gaissmaier, W., & Gigerenzer, G. (2010). Good judgments do not require complex cognition. *Current Directions in Psychological Science*, *11*, 103–121.
Matthews, G., Zeidner, M., & Roberts, R. D. (2002). *Emotional intelligence: Science and myth*. Cambridge, MA: MIT Press.
McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, *306*, 503–507.
Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of will power. *Psychological Review*, *106*, 3–19.
Milkman, K. L., Chugh, D., & Bazerman, M. H. (2009). How can decision making be improved? *Perspectives on Psychological Science*, *4*, 379–383.
Minsky, M. L. (1985). *The society of mind*. New York: Simon and Schuster.
Miyake, A., Friedman, N., Emerson, M. J., & 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.
Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*, 297–326.
Murphy, D., & Stich, S. (2000). Darwin in the madhouse: Evolutionary psychology and the classification of mental disorders. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the human mind: Modularity, language and meta-cognition* (pp. 62–92). Cambridge, England: Cambridge University Press.
Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology*, *57A*, 33–60.
Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (pp. 1–18). New York: Plenum.
Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Midwest Publications.
Oaksford, M., & Chater, N. (1998). An introduction to rational models of cognition. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 1–18). New York: Oxford University Press.
Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford, England: Oxford University Press.
Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior and Organization*, *72*, 147–152.
Over, D. E. (2000). Ecological rationality and its heuristics. *Thinking and Reasoning*, *6*, 182–192.
Over, D. E. (2004). Rationality and the normative/descriptive distinction. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 3–18). Malden, MA: Blackwell.
Parker, A. M., & Fischhoff, B. (2005). Decision-making competence: External validation through an individual differences approach. *Journal of Behavioral Decision Making*, *18*, 1–27.
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.
Plato. (1945). *The republic*. (F. MacDonald Cornford, Trans.). New York: Oxford University Press.
Pollock, J. L. (1995). *Cognitive carpentry*. Cambridge, MA: The MIT Press.
Postman, N. (1988). *Conscientious objections*. New York: Vintage Books.
Prado, J., & Noveck, I. A. (2007). Overcoming perceptual features in logical reasoning: A parametric functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, *19*, 642–657.
Reber, A. S. (1993). *Implicit learning and tacit knowledge*. New York: Oxford University Press.
Rescher, N. (1988). *Rationality: A philosophical inquiry into the nature and rationale of reason*. Oxford, England: Oxford University Press.
Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago, IL: University of Chicago Press.
Rips, L. J. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.
Sá, W., Kelley, C., Ho, C., & Stanovich, K. E. (2005). Thinking about personal theories: Individual differences in the coordination of theory and evidence. *Personality and Individual Differences*, *38*, 1149–1161.
Sá, W., & Stanovich, K. E. (2001). The domain specificity and generality of mental contamination: Accuracy and projection in judgments of mental content. *British Journal of Psychology*, *92*, 281–302.
Sá, W., West, R. F., & 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., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*, *132*, 566–594.
Salthouse, T. A., & Pink, J. E. (2008). Why is working memory related to fluid intelligence? *Psychonomic Bulletin and Review*, *15*, 364–371.
Samuels, R. (2005). The complexity of cognition: Tractability arguments for massive modularity. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind* (pp. 107–121). Oxford, England: Oxford University Press.
Samuels, R. (2009). The magical number two, plus or minus: Dual-process theory as a theory of cognitive kinds. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 129–146). Oxford, England: Oxford University Press.
Samuels, R., & Stich, S. P. (2004). Rationality and psychology. In A. R. Mele & P. Rawling (Eds.), *The Oxford handbook of rationality* (pp. 279–300). Oxford, England: Oxford University Press.
Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Schelling, T. C. (1984). *Choice and consequence: Perspectives of an errant economist*. Cambridge, MA: Harvard University Press.

Schneider, W., & Chein, J. (2003). Controlled and automatic processing: Behavior, theory, and biological processing. *Cognitive Science*, *27*, 525–559.

Schommer-Aikins, M. (2004). Explaining the epistemological belief system: Introducing the embedded systemic model and coordinated research approach. *Educational Psychologist*, *39*, 19–30.

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

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.

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, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 379–396). New York: Cambridge University Press.

Slovic, P., & Peters, E. (2006). Risk perception and affect. *Current Directions in Psychological Science*, *15*, 322–325.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*, 108–131.

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

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, 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), 91–92.

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* (pp. 124–158). 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 & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 291–342). New York: Cambridge University Press.

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

Stanovich, K. E. (2009a). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55–88). Oxford, England: Oxford University Press.

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

Stanovich, K. E. (2010). *Decision making and rationality in the modern world*. New York: Oxford University Press.

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

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*, 342–357.

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

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

Stanovich, K. E., & West, R. F. (1998c). Individual differences in rational thought. *Journal of Experimental Psychology: General*, *127*, 161–188.

Stanovich, K. E., & West, R. F. (1998d). Who uses base rates and P(D/~H)? An analysis of individual differences. *Memory and Cognition*, *26*, 161–179.

Stanovich, K. E., & West, R. F. (1999). Discrepancies between normative and descriptive models of decision making and the understanding/acceptance principle. *Cognitive Psychology*, *38*, 349–385.

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

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

Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking and Reasoning*, *13*, 225–247.

Stanovich, K. E., & West, R. F. (2008a). On the failure of intelligence to predict myside bias and one-sided bias. *Thinking and Reasoning*, *14*, 129–167.

Stanovich, K. E., & West, R. F. (2008b). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*, 672–695.

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

Sternberg, R. J. (2003). *Wisdom, intelligence, and creativity synthesized*. Cambridge, England: Cambridge University Press.

Sternberg, R. J., & Grigorenko, E. L. (1997). Are cognitive styles still in style? *American Psychologist*, *52*, 700–712.

Sternberg, R. J., Grigorenko, E. L., & Zhang, L. (2008). Styles of learning and thinking matter in instruction and assessment. *Perspectives on Psychological Science*, *3*, 486–506.

Stich, S. P. (1990). *The fragmentation of reason*. Cambridge, MA: MIT Press.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220–247.

Strathman, A., Gleicher, F., Boninger, D. S., & Scott Edwards, C. (1994). The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology*, *66*, 742–752.

Taub, G. E., & McGrew, K. S. (2004). A confirmatory factor analysis of Cattell-Horn-Carroll theory and cross-age invariance of the Woodcock-Johnson Tests of Cognitive Abilities III. *School Psychology Quarterly*, *19*, 72–87.

Tetlock, P. E., & Mellers, B. A. (2002). The great rationality debate. *Psychological Science*, *13*, 94–99.
Thagard, P., & Nisbett, R. E. (1983). Rationality and charity. *Philosophy of Science*, *50*, 250–267.
Thaler, R. H., & Shefrin, H. M. (1981). An economic theory of self-control. *Journal of Political Economy*, *89*, 392–406.
Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.
Toates, F. (2005). Evolutionary psychology: Towards a more integrative model. *Biology and Philosophy*, *20*, 305–328.
Toates, F. (2006). A model of the hierarchy of behavior, cognition, and consciousness. *Consciousness and Cognition*, *15*, 75–118.
Toplak, M., Liu, E., Macpherson, R., Toneatto, T., & Stanovich, K. E. (2007). The reasoning skills and thinking dispositions of problem gamblers: A dual-process taxonomy. *Journal of Behavioral Decision Making*, *20*, 103–124.
Toplak, M. E., & 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. & Stanovich, K. E. (2003). Associations between myside bias on an informal reasoning task and amount of post-secondary education. *Applied Cognitive Psychology*, *17*, 851–860.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.
von Neumann, J., & Morgenstern, O. (1944). *The theory of games and economic behavior*. Princeton, NJ: Princeton University Press.
Vranas, P. B. M. (2000). Gigerenzer's normative critique of Kahneman and Tversky. *Cognition*, *76*, 179–193.
Wang, L. (2009). Money and fame: Vividness effects in the National Basketball Association. *Journal of Behavioral Decision Making*, *22*, 20–44.
Wason, P. C., & Evans, J. St. B. T. (1975). Dual processes in reasoning? *Cognition*, *3*, 141–154.
West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology*, *100*, 930–941.
Wilson, T. D. (2002). *Strangers to ourselves*. Cambridge, MA: Harvard University Press.
Wilson, T. D., & Gilbert, D. T. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science*, *14*, 131–134.
Wu, G., Zhang, J., & Gonzalez, R. (2004). Decision under risk. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 399–423). Malden, MA: Blackwell Publishing.
Zeidner, M., & Matthews, G. (2000). Intelligence and personality. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 581–610). New York: Cambridge University Press.
Zelazo, P. D. (2004). The development of conscious control in childhood. *Trends in Cognitive Sciences*, *8*, 12–17.