8

A Framework for Critical Thinking, Rational Thinking, and Intelligence

KEITH E. STANOVICH AND PAULA J. STANOVICH

Critical thinking is highly valued in educational writings if not in practice. Despite a substantial literature on the subject, for many years the area of critical thinking was notorious for its conceptual difficulties. For example, years ago Cuban (1984) lamented that “defining thinking skills, reasoning, critical thought, and problem solving is troublesome to both social scientists and practitioners. Troublesome is a polite word; the area is a conceptual swamp” (p. 676). There has been some progress in elucidating the concept of critical thinking since the time of Cuban’s statement, but we shall argue here that educational theory is on the verge of an even more stunning conceptual advance in the area of critical thinking. Education is beginning to understand the critical thinking concept by relating it to the constructs of intelligence and rational thought. In fact, modern cognitive science provides a coherent framework for understanding the relation between critical thinking, intelligence, and rational thought.

THE FOUNDATIONAL SKILLS OF CRITICAL THINKING

In the critical thinking literature, the ability to evaluate evidence and arguments independently of one’s prior beliefs and opinions is a skill that is strongly emphasized (Baron, 2008; Dole & Sinatra, 1998; Ennis, 1987, 1996; Kuhn, 2005; Lipman, 1991; Paul, 1984, 1987; Ritchhart & Perkins, 2005; Sternberg, 1997, 2001, 2003; Wade & Tavris, 1993). The disposition toward such unbiased reasoning is almost universally viewed as a characteristic of good thinking. For example, Norris and Ennis (1989) argue that one fundamentally important characteristic of critical thinking is the disposition to "reason from starting points with which we disagree without letting the disagreement interfere with reasoning" (p. 12). Zechmeister and Johnson (1992) list as one characteristic of the critical thinker the ability to "accept statements as true even when they don't agree with one's own position" (p. 6). Similarly, Nickerson (1987) stresses that critical thinking entails the ability to recognize "the fallibility of one's own opinions, the probability of bias in those opinions, and the danger of differentially weighting evidence according to personal preferences" (p. 30). The growing literature on informal reasoning likewise emphasizes the importance of detaching one's own beliefs from the process of argument evaluation (Baron, 1995; Klaczynski & Lavallee, 2005; Kuhn, 2001, 2007; Kuhn & Udell, 2007; Macpherson & Stanovich, 2007; Toplak & Stanovich, 2003; Voss, Perkins, & Segal, 1991).

The emphasis on unbiasedness in reasoning has led many theorists to highlight the importance of decontextualization as the foundational skill of critical thinking thought (see Paul, 1984, 1987; Siegel, 1988). For example, Kelley (1990) argues that "the ability to step back from our train of thought . . . . is a virtue because it is the only way to check the results of our thinking, the only way to avoid jumping to conclusions, the only way to stay in touch with the facts" (p. 6). Neimark (1987) lumps the concepts of decentering and decontextualizing under the umbrella term *detachment*. She terms one component of detachment *depersonalizing*: being able to adopt perspectives other than one's own. This aspect of detachment is closely analogous to Piaget's (1926) concept of decentration. Neimark's (1987) second component of detachment—detaching from context—involves breaking the bounds of situational constraint and local context. It is reminiscent of Donaldson's (1978, 1993) concept of disembedding:

> If the intellectual powers are to develop, the child must gain a measure of control over his own thinking and he cannot control it while he remains unaware of it. The attaining of this control means prising thought out of its primitive unconscious embeddedness in the immediacies of living in the world and interacting with other human beings. It means learning to move

> beyond the bounds of human sense. It is on this movement that all the higher intellectual skills depend. (Donaldson, 1978, p. 123)

Neimark (1987) emphasizes how associations built up over time will tend to activate a decision for us automatically and unconsciously if we are not reflective and cannot detach from situational cues. The danger of response patterns that are determined too strongly by overlearned cues is a repeated theme in the heuristics and biases literature of cognitive science (e.g., Arkes, 1991; Evans, 2003, 2006, 2007, 2008; Kahneman, 2003; Kahneman & Frederick, 2002; Stanovich, 2003, 2004, 2009; Wilson & Brekke, 1994). For example, Baron (1994) argues that many departures from consequentialism in decision making are due to inappropriate generalizations. For example, the act–omission distinction is hypothesized to arise because harmful acts are usually more intentional than harmful omissions; but this distinction continues to be made even when there is no difference in intention. In short, to act in consequentialist fashion, the features of the actual context (intention, etc.) must be abstracted and compared componentially. Such decontextualizing cognitive habits represent one line of defense against overlearned associations that might trigger nonnormative responses.

Many tasks in the heuristics and biases branch of the reasoning literature might be said to involve some type of decontextualization skill (Kahneman, 2003; Stanovich, 2003). Tasks are designed to see whether reasoning processes can operate independently of interfering context (world knowledge, prior opinion, vivid examples). One example of such a task is the laboratory paradigm that has been used to study belief bias in syllogistic reasoning. The stimuli in the task put the validity of a syllogism and the facts expressed in the conclusion of the syllogism in conflict. For example, the syllogism "All flowers have petals, roses have petals, therefore roses are flowers" is invalid despite the seeming "rightness" of the conclusion. The inability to decouple prior knowledge from reasoning processes has been termed the *belief bias effect* (Evans, Barston, & Pollard, 1983; Evans, Newstead, Allen, & Pollard, 1994), and it has been the subject of extensive study in the cognitive science literature (De Neys, 2006; Evans, 2002; Evans & Curtis-Holmes, 2005; Evans & Feeney, 2004; Garnham & Oakhill, 2005; Klauer, Musch, & Naumer, 2000).

Belief bias has also been revealed in paradigms where participants must evaluate the quality of empirical evidence in a manner not contaminated by their prior opinions on the issue in question. In several studies, Klaczynski and colleagues (Klaczynski, 1997; Klaczynski & Lavallee, 2005; Klaczynski & Robinson, 2000) presented participants with flawed hypothetical experiments leading to conclusions that were either consistent or inconsistent with prior positions and opinions. Participants then critiqued the flaws in the experiments (which were most often badly flawed). Participants found many more flaws when the experiment's conclusions were inconsistent with their prior opinions than when the experiment's conclusions were consistent with their prior opinions and beliefs (see also Macpherson & Stanovich, 2007).

CRITICAL THINKING IN THE SERVICE OF RATIONAL THOUGHT

Like the study of wisdom (Sternberg, 2001, 2003; Sternberg & Jordan, 2005), the study of critical thinking is a normative/evaluative endeavor. Specifically, if one's goal is to *aid* people in their thinking, then it is essential that one have some way of *evaluating* thinking. For example, in the current educational literature, teachers are constantly exhorted to "teach children how to think," or to foster "critical thinking" and "creative problem solving." However, the problem here is that "thinking" is not a domain of knowledge. As Baron (1993) notes,

> We teach Latin or calculus because students do not already know how to speak Latin or find integrals. But, by any reasonable description of thinking, students already know how to think, and the problem is that they do not do it as effectively as they might. (p. 199)

Thus the admonition to educators to "teach thinking skills" and foster "critical thinking" contains implicit evaluative assumptions. The children *already* think. Educators are charged with getting them to think *better* (Adams, 1993). This, of course, implies a normative model of what we mean by better thinking (Baron, 1993, 2008).

A somewhat analogous issue arises when thinking dispositions are discussed in the educational literature of critical thinking. Why do we want people to think in an actively open-minded fashion? Why do we want to foster multiplist and evaluative thinking (Kuhn, 1993, 2001, 2005; Kuhn & Udell, 2007) rather than absolutist thinking? Why do we want people to be reflective? It can be argued that the superordinate goal we are actually trying to foster is that of rationality (Stanovich, 2004, 2009). That is, much of what educators are ultimately concerned about is rational thought in both the epistemic sense and the practical sense. We value certain thinking dispositions because we think that they will at least aid in bringing belief in line with the world and in achieving our goals. By a parallel argument, we could equally well claim that the superordinate goal is to educate for wisdom (Sternberg, 2001, 2002, 2003).

We can see that it is rationality, and not critical thinking per se, that is the ultimate goal of education by conducting some simple thought experiments or imaginative hypotheticals. For example, we could imagine a person with excellent epistemic rationality (his or her degree of confidence in propositions being well calibrated to the available evidence relevant to the proposition) and optimal practical rationality (the person optimally satisfies desires) who was *not* actively open-minded—that is, who was not a good critical thinker under standard assumptions. Of course we would still want to mold such an individual's dispositions in the direction of open-mindedness for the sake of society as a whole. But the essential point for the present discussion is that, from a purely *individual* perspective, we would now be hard pressed to find reasons why we would *want* to change such a person's thinking dispositions—whatever they were—if they had led to rational thought and action in the past.

In short, a large part of the rationale for educational interventions to change thinking dispositions derives from a tacit assumption that actively open-minded critical-thinking dispositions make the individual a more rational person—or as Sternberg (2001, 2005) argues, a wiser, less foolish person. Thus, the normative justification for fostering critical thought is that it is the foundation of rational thought. The thinking dispositions associated with critical thinking must be fostered because they make students more rational. Our view is consistent with that of many other theorists who have moved toward conceptualizing critical thinking as a subspecies of rational thinking or at least as closely related to rational thinking (Kuhn, 2005; Moshman, 2004, 2005, 2010; Reyna, 2004; Siegel, 1988, 1997).

The grounding of critical thinking within the concept of rationality in this manner has many conceptual advantages. First, the concept of rationality is deeply intertwined with the data and theory of modern cognitive science (see LeBoeuf & Shafir, 2005; Over, 2004; Samuels & Stich, 2004; Stanovich, 2004, 2009) in a way that the concept of critical thinking is not. Additionally, as we demonstrate below, theory in cognitive science differentiates rationality from intelligence and explains why rationality and intelligence often dissociate. This finding and its explanation confirm the long-standing belief in education that intelligence does not guarantee critical thinking.

RATIONAL THOUGHT AND ITS OPERATIONALIZATIONS IN COGNITIVE SCIENCE

Cognitive scientists recognize two types of rationality: instrumental and epistemic. The simplest definition of instrumental rationality is 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 is one in which a person chooses options based on which option has the largest expected utility (see Baron, 2008; Dawes, 1998; Hastie & Dawes, 2001; Wu, Zhang, & Gonzalez, 2004).

The other aspect of rationality studied by cognitive scientists is termed *epistemic rationality*. This aspect of rationality concerns how well beliefs map onto the actual structure of the world. Epistemic rationality is sometimes called theoretical rationality or evidential rationality (see Audi, 1993, 2001; Foley, 1987; Harman, 1995; Manktelow, 2004; Over, 2004). Instrumental and epistemic rationality are related. The aspect of beliefs that enter into instrumental calculations (that is, tacit calculations) are the probabilities of states of affairs in the world.

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), they are behaving as if they were maximizing utility—they are acting to get what they most want (Edwards, 1954; 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 has been the logic of the research program on heuristics and biases inaugurated in the much-cited studies of Kahneman and Tversky (1972, 1973, 1979; Tversky & Kahneman, 1974, 1981, 1983, 1986).

Researchers in the tradition of heuristics and biases have demonstrated in a host of empirical studies that people violate many of the strictures of rationality and that the magnitude of these violations can be measured experimentally. For example, people display confirmation bias, test hypotheses inefficiently, display preference inconsistencies, do not properly calibrate degrees of belief, overproject their own opinions onto others, combine probabilities incoherently, and allow prior knowledge to become implicated in deductive reasoning (for summaries of the large literature, see Baron, 2008; Evans, 1989, 2007; Gilovich, Griffin, & Kahneman, 2002; Kahneman & Tversky, 2000; Shafir & LeBoeuf, 2002; Stanovich, 1999, 2004, 2009). These are caused by many well-known cognitive biases: base-rate neglect, framing effects, representativeness biases, anchoring biases, availability bias, outcome bias, and vividness effects, to name just a few. Degrees of rationality can be assessed in terms of the number and severity of such cognitive biases that individuals display. Failure to display a bias becomes a measure of rational thought.

In an attempt to understand how these various errors in rational thinking originate, investigators working in the tradition of heuristics and biases have been inexorably drawn to dual-process models of cognitive architecture. Recently, in an attempt to extend these models into the domain of individual differences, Stanovich (2009) has proposed a triprocess distinction that both explains errors in heuristics and biases tasks and, even more importantly, elucidates the relation between rationality and intelligence.

DUAL-PROCESS MODELS OF COGNITION

Virtually all attempts to classify heuristics and biases tasks end up utilizing a dual-process framework because most of the tasks in the literature on heuristics and biases were deliberately designed to pit a heuristically triggered response against a normative response generated by the analytic system. As Kahneman (2000) notes, “Tversky and I always thought of the heuristics and biases approach as a two-process theory” (p. 682). Since Kahneman and Tversky launched the heuristics and biases approach in the 1970s, a wealth of evidence has accumulated in support of the dual-process approach. Evidence from cognitive neuroscience and cognitive psychology converges on the conclusion that mental functioning can be characterized by two different types of cognition having somewhat different functions and different strengths and weaknesses (Brainerd & Reyna, 2001; Evans, 2003, 2008, 2009; Evans & Over, 1996, 2004; Feldman Barrett, Tugade, & Engle, 2004; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Kahneman & Frederick, 2002, 2005; Lieberman, 2003; McClure, Laibson, Loewenstein, & Cohen, 2004; Metcalfe & Mischel, 1999; Sloman, 1996, 2002; Stanovich, 1999, 2004).

There are many such theories (over 20 dual-process theories are presented in a table in Stanovich, 2004) and they have some subtle differences, but they are similar in that all distinguish autonomous from nonautonomous processing. The two types of processing were termed systems in earlier writings, but theorists have been moving toward more atheoretical characterizations; we therefore follow Evans (2009) in using the terms *type 1* and *type 2 processing*.

The defining feature of type 1 processing is its autonomy. Type 1 processes are termed autonomous because (a) their execution is rapid, (b) their execution is mandatory when the triggering stimuli are encountered, (c) they do not put a heavy load on central processing capacity (i.e., they do not require conscious attention), (d) they do not depend on input from high-level control systems, and (e) they can operate in parallel without interfering with each other or with type 2 processing. Type 1 processing 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 (see Evans, 2007, 2008; Stanovich, 2004, 2009).

Type 2 processing contrasts with type 1 processing on each of the critical properties that define the latter. Type 2 processing is relatively slow and computationally expensive—it is the focus of our awareness. Many type 1 processes can operate at once in parallel, but only one (or a very few) type 2 thoughts can be executing at once—type 2 processing is thus serial processing. Type 2 processing is often language-based.

One of the most critical functions of type 2 processing is to override type 1 processing. 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. In order to override type 1 processing, type 2 processing must display at least two (possibly 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 recent work on executive functioning (Hasher, Lustig, & Zacks, 2007; Miyake, Friedman, Emerson, & Witzki, 2000; Zelazo, 2004).

But the ability to suppress type 1 processing gets the job only half done. Suppressing one response is not helpful unless a better response is 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 unique to type 2 processing (Evans, 2007; Evans & Over, 2004; Kahneman & Tversky, 1982; Nichols & Stich, 2003). When we reason hypothetically, we create temporary models of the world and test out actions (or alternative causes) in that simulated world. In order to reason hypothetically we must, however, have one critical cognitive capability—the ability to distinguish our representations of the real world from representations of imaginary situations. For example, in considering an alternative goal state different from the one we currently have, we must be able to represent our current goal and the alternative goal and to keep straight which is which. Likewise, we need to be able to differentiate the representation of an action about to be taken from representations of potential *alternative* actions we are considering. But the latter must not infect the former while the mental simulation is being carried out. Thus, several years ago 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 do any other type 2 operation. In fact, decoupling operations might well be a major contributor to a distinctive type 2 property—its seriality.

Cognitive decoupling must take place when an individual engages in a simulation of alternative worlds in order to solve a problem. Problem-solving tasks that necessitate fully disjunctive reasoning (see Johnson-Laird, 2006; Shafir, 1994) provide examples of the situations that require fully decoupled simulation. Fully disjunctive reasoning involves considering all possible states of the world in deciding among options or in choosing a problem solution in a reasoning task. Consider the following problem, taken from the work of Levesque (1986, 1989) and studied by our research group (see Toplak & Stanovich, 2002): Jack is looking at

Anne but Anne is looking at George. Jack is married but George is not. Is a married person looking at an unmarried person?

A. Yes
B. No
C. Cannot be determined

The vast majority of people answer C (cannot be determined) when in fact the correct answer to this problem is A (yes). To answer correctly, both possibilities for Anne's marital status (married and unmarried) must be considered to determine whether a conclusion can be drawn. If Anne is married, then the answer is yes because she would be looking at George, who is unmarried. If Anne is not married, then the answer is still yes because Jack, who is married, would be looking at Anne. Considering all the possibilities (the fully disjunctive reasoning strategy) reveals that a married person is looking at an unmarried person whether Anne is married or not. The fact that the problem does not *reveal* whether Anne is married suggests to people that nothing can be determined. Many people make the easiest (incorrect) inference from the information given and do not proceed with the more difficult (but correct) inference that follows from fully disjunctive reasoning.

Not all type 2 processing represents fully explicit cognitive simulation, however. Or, to put it another way: all hypothetical thinking involves type 2 processing (Evans & Over, 2004) but not all type 2 processing involves hypothetical thinking. What has been termed serial associative cognition (Stanovich, 2009) represents this latter category. It can be understood by considering a discussion of the four-card selection task in a theoretical paper on dual processes by Evans (2006; see also Evans & Over, 2004). In Wason's (1966) four-card selection task, the participant is told the following:

> Each of the boxes below represents a card lying on a table. Each one of the cards has a letter on one side and a number on the other side. Here is a rule: If a card has a vowel on its letter side, then it has an even number on its number side. As you can see, two of the cards are letter-side up, and two of the cards are number-side up. Your task is to decide which card or cards must be turned over in order to find out whether the rule is true or false. Indicate which cards must be turned over.

The participant chooses from four cards labeled K, A, 8, and 5 (corresponding to not-P, P, Q, and not-Q). The correct answer is to pick the

A and the 5 (P and not-Q), but the most common answer is to pick the A and 8 (P and Q)—the so-called matching response.

Evans (2006) points out that the previous emphasis on the matching bias evident in the task (Evans, 1972, 1998; Evans & Lynch, 1973) might have led some investigators to infer that type 2 processing does not occur. In fact, matching bias might be viewed as just one of several such suggestions in the literature that much thinking during the task is type 1 processing (see Hardman, 1998; Margolis, 1987; Stanovich & West, 1998a; Tweney & Yachanin, 1985). In contrast, however, Evans (2006) presents evidence indicating that type 2 processing may be going on during the task—even on the part of the majority who do not give the normatively correct response but instead give the PQ response.

In fact, type 2 processing is occurring in this task, but it is not full-blown cognitive simulation of alternative world models. It is thinking of a shallower type—serial associative cognition. Serial associative cognition is not rapid and parallel, such as type 1 processes, but is nonetheless rather inflexibly locked into an associative mode that takes as its starting point a model of the world that is *given* to the subject. For example, Evans and Over (2004) note that in the studies of verbal protocols, when participants made an incorrect choice, they referred to the hidden sides of the cards they were going to pick, but referred only to verification when they did so. Thus, the evidence suggests that people accept the rule as given, assume it is true, and simply describe how they would go about verifying it. The fact that they refer to hidden sides does not mean that they have constructed any alternative model of the situation beyond what was given to them by the experimenter and their own assumption that the rule is true. They then reason from this single focal model—systematically generating associations from this focal model but never constructing another model of the situation. This is why the central characteristic of serial associative cognition is that it displays a *focal bias.*

One way in which to contextualize the idea of focal bias is as the second stage in a framework for thinking about human information processing that dates to the mid 1970s—the idea of humans as cognitive misers (Dawes, 1976; Taylor, 1981; Tversky & Kahneman, 1974). There are, in fact, two aspects of cognitive miserliness. Dual-process theory has heretofore highlighted only rule 1 of the cognitive miser: default to type 1 processing whenever possible. But defaulting to type 1 processing is not always possible—particularly in novel situations where there are no stimuli available to trigger domain-specific evolutionary modules. Type 2 processing procedures will be necessary, but a cognitive miser default is operating even there. Rule 2 of the cognitive miser is that when type 1 processing will not yield a solution, default to serial associative cognition with a focal bias (*not* fully decoupled cognitive simulation).

So the focal model that will dominate processing—the only model that serial associative cognition deals with—is the most easily constructed model. The most easily constructed model: tends to represent only one state of affairs; it accepts what is directly presented and models what is presented as true; it ignores moderating factors—probably because taking account of those factors would necessitate modeling several alternative worlds and this is just what a focal processing allows us to avoid. And finally, given the voluminous literature in cognitive science on belief bias and the informal reasoning literature on myside bias, the easiest models to represent clearly appear to be those closest to what a person already believes and has modeled previously (e.g., Evans & Feeney, 2004; Stanovich & West, 2007, 2008a). Thus, serial associative cognition is defined by its reliance on a single focal model that triggers all subsequent thought.

THREE KINDS OF MINDS

In 1996 philosopher Daniel Dennett wrote a book about how aspects of the human mind were like the minds of other animals and how other aspects were not. He titled the book *Kinds of Minds* to suggest that within the brain of humans are control systems of very different types—different kinds of minds. We are going to make here a distinction between aspects of type 2 processing that can be introduced through a Dennett-type example of different types of explanation. Imagine two different stories involving a woman walking on a cliff. The stories are all sad—the woman dies in each. The purpose of this exercise is to get us to think about how we explain the death in each story. In incident A, the woman is walking on a cliffside by the ocean and goes to step on a large rock, but the rock is not a rock at all. Instead, it is actually the side of a crevice and she falls down the crevice and dies. In incident B, the woman attempts suicide by jumping off a cliff and dies when she is crushed on the rocks below.

In both cases, at the most basic level, when we ask ourselves for an explanation of why the woman died, we might say that the answer is the same. The same laws of physics that are in operation in incident A (the gravitational laws that describe why the woman will be crushed upon impact) are also operative in incident B. However, we feel that the laws of gravity and force somehow do not provide a complete explanation of what has happened in incident B. This feeling is correct. The examples each call for a different level of explanation if we wish to zero in on the *essential* cause of death.

In analyzing incident A, a psychologist would be prone to say that in processing a stimulus (the crevice that looked somewhat like a rock) the woman's information processing system malfunctioned—sending the wrong information to response decision mechanisms, which then resulted in a disastrous motor response. Cognitive scientists refer to this level of analysis as the algorithmic level (Anderson, 1990; Marr, 1982; Stanovich, 1999). In the realm of machine intelligence, this would be the level of the instructions in the abstract computer language used to program the computer (FORTRAN, COBOL, etc.). The cognitive psychologist works largely at this level by showing that human performance can be explained by positing certain information processing mechanisms in the brain (input coding mechanisms, perceptual registration mechanisms, storage systems of short- and long-term memory, etc.). For example, a simple letter pronunciation task might entail encoding the letter, storing it in short-term memory, and comparing it with information stored in long-term memory and, if a match occurs, making a response decision and executing a motor response. In the case of the woman in incident A, the algorithmic level is the right level to explain her unfortunate demise. Her perceptual registration and classification mechanisms malfunctioned by providing incorrect information to response decision mechanisms, causing her to step into the crevice.

Incident B, on the other hand, does not involve such an information processing error at the algorithmic level. The woman's perceptual apparatus accurately recognized the edge of the cliff and her motor command centers accurately programmed her body to jump off the cliff. The computational processes posited at the algorithmic level of analysis executed quite perfectly. No error at this level of analysis explains why the woman is dead in incident B. Instead, this woman died because of her overall goals and how these goals interacted with her beliefs about the world in which she lived.

In the spirit of Dennett's book, Stanovich (2009) termed the part of the mind that carries out type 1 processing the *autonomous mind*. Different kinds of type 2 processing are defined by incidents A and B in the imaginary scenarios. In the terms of Stanovich (2009), the woman in incident A had a problem with the algorithmic mind and the woman in incident B had a problem with the reflective mind. This terminology captures the fact that we turn to an analysis of goals, desires, and beliefs to understand a case such as B. The algorithmic level provides an incomplete explanation of behavior in cases like incident B because it provides an information processing explanation of how the brain is carrying out a particular task (in this case, jumping off of a cliff) but no explanation of *why* the brain is carrying out this particular task. We turn to the level of the reflective mind, where we ask questions about the *goals* of the system's computations (*what* the system is attempting to compute and *why*). In short, the reflective mind is concerned with the goals of the system, beliefs relevant to those goals, and the choice of action that is optimal given the system's goals and beliefs. It is only at the level of the reflective mind that issues of rationality come into play. Importantly, the algorithmic mind can be evaluated in terms of efficiency but not rationality.

This concern for the efficiency of information processing as opposed to its rationality is mirrored in the status of intelligence tests. They are measures of efficiency but not rationality—a point made clear by considering a distinction that is very old in the field of psychometrics. Psychometricians have long distinguished typical performance situations from optimal (sometimes termed maximal) performance situations (see Ackerman, 1994, 1996; Ackerman & Heggestad, 1997; Ackerman & Kanfer, 2004; see also, Cronbach, 1949; Matthews, Zeidner, & Roberts, 2002). 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 and is told how to do so. Thus, optimal performance measures examine questions of 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, Petty, & Feinstein 1996; Kruglanski & Webster, 1996; Norris & Ennis, 1989; Schommer-Aikins, 2004; Stanovich, 1999, 2009; Sternberg, 2003; Sternberg & Grigorenko, 1997; Strathman, Gleicher, Boninger, & Scott Edwards, 1994).

The literature on these types of thinking dispositions is vast and our purpose is not to review that literature here. It is only necessary to note that the types of cognitive propensities that these thinking disposition measures reflect are the tendency to collect information before making up one's mind, to seek various points of view before coming to a conclusion, to think extensively about a problem before responding, to calibrate the degree of strength of one's opinion to the degree of evidence available, to think about future consequences before taking action, to explicitly weigh pluses and minuses of situations before making a decision, and to seek nuance and avoid absolutism. In short, individual differences in thinking dispositions include assessing variation in people's goal management, epistemic values, and epistemic self-regulation—differences in the operation of reflective mind. They are all 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, the tendency to change beliefs in the face of contrary evidence, or 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.

The algorithmic mind, assessed on actual IQ tests, is relevant in determining what happened in the case A above, but it does not provide sufficient explanation of case B. To understand what happened in case B, we need to know about more than the woman's processes of memory and speed of pattern recognition. We need to know what her goals were and what she believed about the world. And one of the most pressing things we want to know about the woman in case B is whether there was some sense in her jumping off the cliff. We do not want to know whether she threw herself off with the greatest efficiency possible (an algorithmic-level question). We want to know whether it was *rational* for her to jump.

A TRIPARTITE MODEL OF MIND AND INDIVIDUAL DIFFERENCES

We have now bifurcated the notion of type 2 processing into two different things—the reflective mind and the algorithmic mind. Previous dual-process views tended to ignore individual differences and hence to miss critical differences in type 2 processing. Figure 8.1 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. The figure identifies variation in fluid intelligence with individual differences in the efficiency of processing of the algorithmic mind. In contrast, thinking dispositions index individual differences in the reflective mind. The reflective and algorithmic minds are characterized by continuous individual differences. Continuous individual differences in the autonomous mind are few. Disruptions to the autonomous mind often reflect damage to cognitive modules, which results in very discontinuous cognitive dysfunction such as autism or the agnosias and alexias (Anderson, 2005; Bermudez, 2001; Murphy & Stich, 2000).

Figure 8.1 highlights an important sense in which rationality is a more encompassing construct than intelligence. 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 allows the correct beliefs to be fixed and the correct actions to be taken. Thus individual differences in rational thought and action can arise because of individual differences in intelligence (the algorithmic mind) or because of individual differences in thinking dispositions (the reflective mind). To put it simply, the concept of ratio-

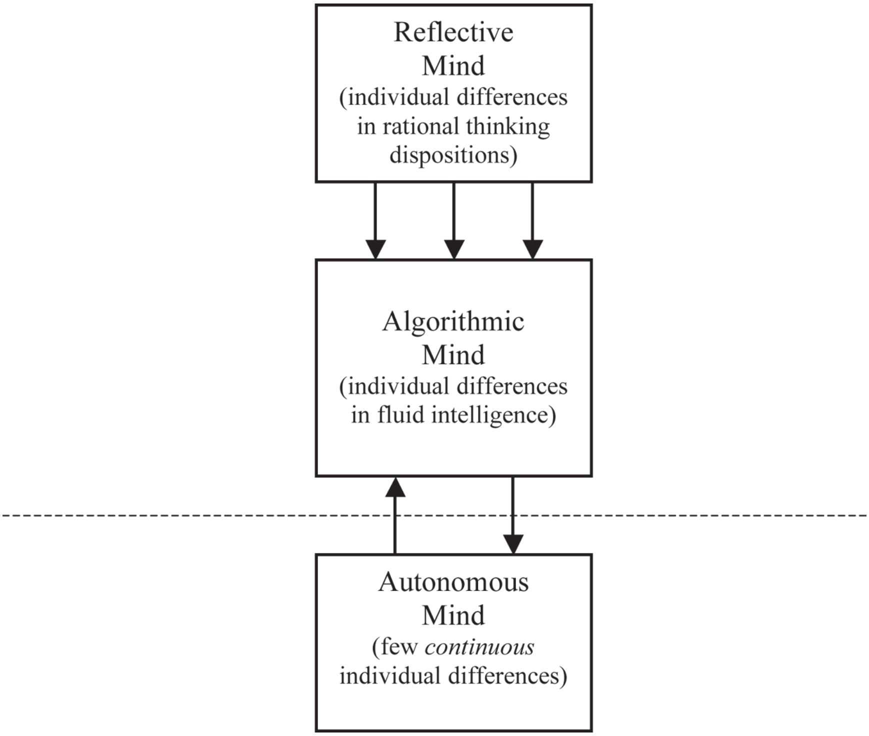


Figure 8.1 Individual differences in the tripartite model.

nality encompasses two things—thinking dispositions of the reflective mind and algorithmic-level efficiency—whereas the concept of intelligence, at least as it is commonly operationalized, is largely confined to algorithmic-level efficiency.

The conceptualization in Figure 8.1 has two great advantages. First, it conceptualizes intelligence in terms of what intelligence tests actually measure. That is, all current tests assess various aspects of algorithmic efficiency. But that is all they assess. None attempt to measure directly an aspect of epistemic or instrumental rationality, nor do they examine any thinking dispositions that relate to rationality. It seems perverse to define intelligence as including rationality when no existing IQ test measures any such thing!

The best-known indicators of cognitive functioning—intelligence and cognitive ability tests—do not assess a critical aspect of thinking, the ability to think rationally. To think rationally means adopting appropriate goals, taking the appropriate action given one's goals and beliefs, and holding beliefs that are commensurate with available evidence. Standard intelligence tests do not assess such functions (Perkins, 1995, 2002; Stanovich, 2002, 2009; Sternberg, 2003, 2006). For example, although intelligence tests do assess the ability to focus on an immediate goal in the face of distraction, they do not assess at all whether a person has the tendency to develop goals that are rational in the first place. Likewise, intelligence tests provide good measures of how well a person can hold beliefs in short-term memory and manipulate those beliefs, but they do not assess at all whether a person has the tendency to *form* beliefs rationally when presented with evidence. And again, similarly, intelligence tests give good measures of how efficiently a person processes information that has been provided, but they do not at all assess whether the person is a *critical assessor* of information as it is gathered in the natural environment.

It is clear from Figure 8.1 why rationality and intelligence can become dissociated. 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. 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, Macpherson, Toplak, West, & Stanovich, 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 found that rational thinking dispositions will predict variance in these tasks after the effects of general intelligence have been controlled (Bruine de Bruin, Parker, & Fischhoff, 2007; 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).

Measures of thinking dispositions tell us about the individual's goals and epistemic values—and they index broad tendencies of pragmatic and epistemic self-regulation at a high level of cognitive control. The empirical studies cited indicate that these different types of cognitive predictors are tapping separable variance, and the reason that this is to be expected is because cognitive capacity measures such as intelligence and thinking dispositions map on to different levels in the tripartite model.

The functions of the different levels of control are illustrated more completely in Figure 8.2. There, it is clear that the override capacity itself is a property of the algorithmic mind; 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 8.2, it is represented by 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 we have presented evidence indicating that the two functions are indexed by different types of individual differences: the ability to sustain the inhibition of the type 1 response is indexed by measures of fluid intelligence, and the tendency to initiate override operations is indexed by thinking dispositions such as reflectiveness and need for cognition.

Figure 8.2 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 making the override worthwhile. Figure 8.2 explicitly represents the simulation function as well as the fact that the call to initiate simulation originates in the reflective mind. The decoupling operation (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 distinct 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.

The model in Figure 8.2 defines a third critical function for the algorithmic mind in addition to type 1 processing override and enabling

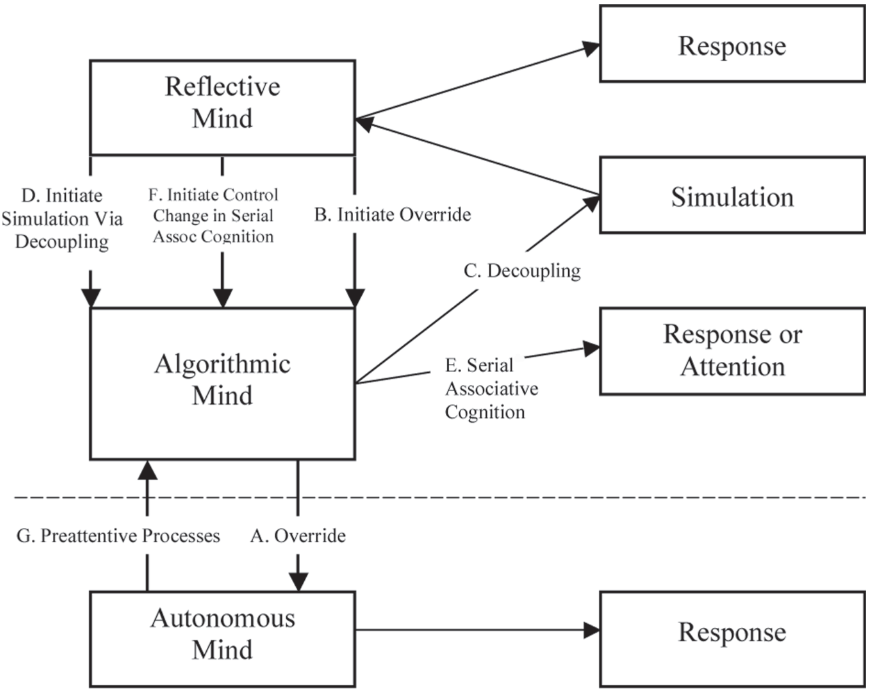


Figure 8.2 A more complete model of the tripartite structure.

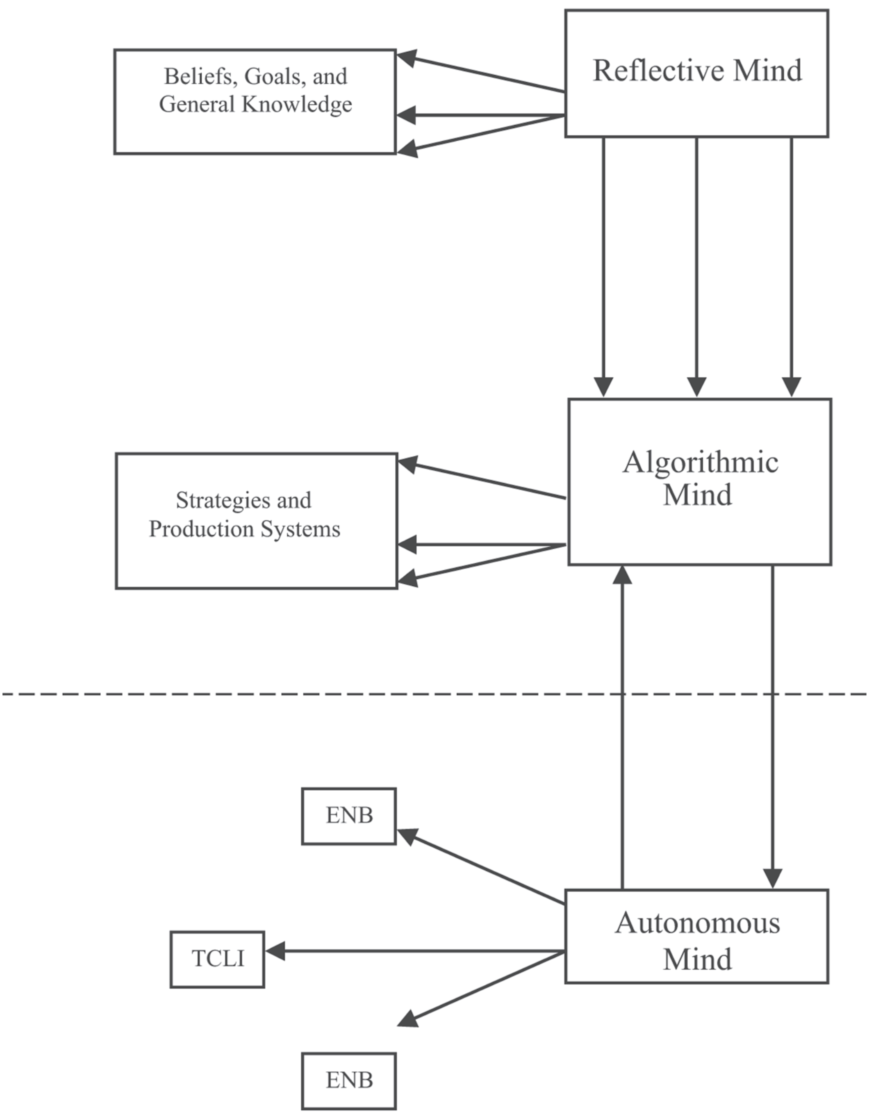
simulation via decoupling. This third function is to sustain the serial associative cognition discussed previously (arrow labeled E). This function is there to remind us that not all type 2 processing involves strongly decoupled cognitive simulation. There are types of slow, serial cognition that do not involve simulating alternative worlds and exploring them exhaustively. The figure thus identifies a third function of the reflective mind—initiating an interrupt of serial associative cognition (arrow F). This interrupt signal alters the next step in a serial associative sequence that would otherwise direct thought. This interrupt signal might have a variety of outcomes. It might stop serial associative cognition altogether in order to initiate a comprehensive simulation (arrow C). Alternatively, it might start a new serial associative chain (arrow E) from a different starting point by altering the temporary focal model that is the source of a new associative chain. Finally, the algorithmic mind receive inputs from the computations of the autonomous mind (arrow G) via so-called preattentive processes (Evans, 2006, 2007, 2008, 2009).

THE IMPORTANCE OF MINDWARE

The term *mindware* was coined by David 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. Perkins uses the term to stress the analogy to software in the brain/computer analogy. Each of the levels in the tripartite model of mind has to access knowledge to carry out its operations, as illustrated in Figure 8.3. As the figure indicates, the reflective mind accesses not only general knowledge structures but, importantly, also 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 evolutionarily compiled encapsulated knowledge bases but also information that has become tightly compiled and available to the autonomous mind because of overlearning and practice.

It is important to note that Figure 8.3 displays 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 8.2). The mindware available for retrieval, particularly that available to the reflective mind, is in part the product of past learning experiences. And here we have a direct link to the Cattell/Horn/Carroll theory of intelligence (Carroll, 1993; Cattell, 1963, 1998; Horn & Cattell, 1967), sometimes termed the fluid/crystallized (or Gf/Gc) theory. The theory posits that tests of mental ability tap a small number of broad factors, of which two are dominant. Fluid intelligence (Gf) reflects reasoning abilities operating across a variety of domains, including novel ones. It is measured by tests of abstract thinking 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).

Knowledge Structures



ENB = Encapsulated Knowledge Base
TCLI = Tightly Compiled Learned Information

Figure 8.3 Knowledge structures in the tripartite model.

The knowledge structures available for retrieval by the reflective mind represent Gc (intelligence as knowledge). Recall that Gf (intelligence as process), is already represented in the Figure 8.2. It is the general computational power of the algorithmic mind—importantly exemplified by the ability to sustain cognitive decoupling.

Because the Gf/Gc theory is one of the more comprehensive theories of intelligence available that has extensive scientific validation, it is thus important to see how both of its major components miss critical aspects of rational thought. Gf 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 Figure 8.2) and to initiate simulation activities (arrow D in Figure 8.2) 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 (it clusters in the domains of probabilistic reasoning, causal reasoning, and scientific reasoning; see Stanovich, 2009). 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.

THE REQUIREMENTS OF RATIONAL THINKING

With this discussion of mindware, we have established that rationality requires three different classes of mental characteristic. First, algorithmic-level cognitive capacity is needed in order that override and simulation activities can be sustained. Second, the reflective mind must be characterized by the tendency to initiate the override of suboptimal responses generated by the autonomous mind and to initiate simulation activities that will result in a better response. Finally, the mindware that allows the computation of rational responses needs to be available and accessible during simulation activities. Intelligence tests assess only the first of these three characteristics that determine rational thought and action. As measures of rational thinking, they are radically incomplete.

Problems in rational thinking arise when the 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, helps to explain one data trend that has been uncovered (Stanovich & West, 2007, 2008a)—that many rational thinking problems are markedly independent of cognitive ability. Ever since Charles Spearman inaugurated the modern period of intelligence research in 1904, what he then termed positive manifold has been the ubiquitous finding: that intelligence indicators have correlated with a plethora of cognitive abilities and thinking skills that are almost too large to enumerate. This is why, among psychologists and among the lay public alike, assessments of intelligence 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 mental traits outside of the cognitive domain, but these critiques concede too much. The tests miss critical thinking processes that are themselves cognitive—the numerous components of rational thinking.

Although the tests fail to assess rational thinking directly, it could be argued that the processes that are tapped by IQ tests largely overlap with variation in rational thinking ability. It is just this conjecture that research has contradicted. Consider the Levesque (1986, 1989) problem (e.g., "Jack is looking at Anne but Anne is looking at George") discussed above. The subjects who answer this problem correctly are no higher in intelligence than those who do not, at least in a sample of university students (see Toplak & Stanovich, 2002).

Most people can carry out fully disjunctive reasoning when they are explicitly *told* that it is necessary. But it is also true that most do not automatically do so. We might expect high-IQ individuals to excel at disjunctive reasoning when they know it is required for successful task performance. But the higher-IQ people in our sample were only slightly more likely to *spontaneously* adopt this type of processing in situations that do not explicitly require it. Note that the instructions in Levesque's

Anne problem do not cue the subject to engage in fully disjunctive reasoning. If *told* to reason through all of the alternatives, the subjects of higher intelligence probably would have done so more efficiently. However, without that instruction, they defaulted to computationally simple cognition in solving problems—they were cognitive misers like everyone else (see Stanovich, 2009). Intelligence and the tendency toward *spontaneous* disjunctive reasoning can be quite unrelated.

This tendency to process information incompletely has been a major theme throughout the past 30 years of research in psychology and cognitive science (Dawes, 1976; Taylor, 1981; Tversky & Kahneman, 1974). For example, Kahneman and Frederick (2002) have shown how people engage in attribute substitution—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 salience for the more effortful retrieval of relevant facts. 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.

A simple example of miserly processing is discussed by Kahneman and Frederick (2002). They describe a simple experiment in which people were asked to consider the following puzzle: "A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?"

Many people offer the response that first comes to mind—10¢–without thinking further and realizing that this cannot be right. The bat would then have to cost $1.10 and the total cost would be $1.20 rather than the required $1.10. People often do not think deeply enough to make this simple correction though, and many students at very selective universities will answer incorrectly and move on to the next problem before realizing that their shallow processing has led them to make an error. Frederick (2005) has found that large numbers of highly selected students at MIT, Princeton, and Harvard, when given this and other similar problems, are cognitive misers like the rest of us. The correlation between intelligence and a set of similar items is quite modest, in the range of .40 to .50 (Gilhooly & Murphy, 2005).

Many other biases of the cognitive miser show correlations no greater than those shown in the Frederick bat-and-ball problem. In fact, some cognitive biases are almost totally dissociated from intelligence. 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 not just because of miserly processing tendencies but also 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. Correlations with cognitive ability have been found to be roughly (in absolute magnitude) in the range of .20 to .35 for probabilistic reasoning tasks and scientific reasoning tasks measuring a variety of rational principles (Bruine de Bruin, et al., 2007; Kokis et al., 2002; Parker & Fischhoff, 2005; Sá, West, & Stanovich, 1999; Stanovich & West, 1997, 1998b, 1998c, 1998d, 1999, 2000; 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 of rational thought 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.

Because the tasks used in this research are so various, we summarize some of this evidence by presenting Tables 8.1 and 8.2. Table 8.1 presents a sampling of rational thinking tasks—each task illustrating an important principle of rational thought—that have shown virtually no relation with intelligence in university samples. Table 8.2 presents a selection of effects and biases that show correlations in the .20 to .35 range.

Rationality is a multifarious concept—not a single mental quality. It requires various thinking dispositions that act to trump a variety of miserly information processing tendencies. It depends on the presence of various knowledge bases related to probabilistic thinking and scientific thinking. It depends on avoiding contaminated mindware that fosters

Table 8.1

TASKS THAT FAIL TO SHOW ASSOCIATIONS WITH COGNITIVE ABILITY

Noncausal base-rate usage (Stanovich & West, 1998c, 1999, 2008)
Conjunction fallacy between subjects (Stanovich & West, 2008)
Framing between subjects (Stanovich & West, 2008)
Anchoring effect (Stanovich & West, 2008)
Evaluability less is more effect (Stanovich & West, 2008)
Proportion dominance effect (Stanovich & West, 2008)
Sunk cost effect (Stanovich & West, 2008; Parker & Fischhoff, 2005)
Risk/benefit confounding (Stanovich & West, 2008)
Omission bias (Stanovich & West, 2008)
Perspective bias (Stanovich & West, 2008)
Certainty effect (Stanovich & West, 2008)
WTP/WTA difference (Stanovich & West, 2008)
My-side bias between and within S (Stanovich & West, 2007, 2008)
Newcomb's problem (Stanovich & West, 1999; Toplak & Stanovich, 2002)

Table 8.2

TASKS THAT SHOW .20–35 ASSOCIATIONS WITH COGNITIVE ABILITY

Causal base-rate usage (Stanovich & West, 1998c, 1998d)
Outcome bias (Stanovich & West, 1998c, 2008)
Framing within subjects
(Frederick, 2005; Parker & Fischhoff, 2005; Stanovich & West, 1998b, 1999)
Denominator neglect (Stanovich & West, 2008; Kokis et al., 2002)
Probability matching (Stanovich & West, 2008; West & Stanovich, 2003)
Hindsight bias (Stanovich & West, 1998c)
Ignoring P(D/NH) (Stanovich & West, 1998d, 1999)
Covariation detection (Stanovich & West, 1998c, 1998d; Sá et al., 1999)
Belief bias in syllogistic reasoning (Stanovich & West, 1998c, 2008)
Belief bias in modus ponens (Stanovich & West, 2008)
Informal argument evaluation (Stanovich & West, 1997, 2008)
Four-card selection task (Stanovich & West, 1998a, 2008)
EV maximization in gambles (Frederick, 2005; Benjamin & Shapiro, 2005)

irrational thought and behavior for its own ends (Blackmore, 1999; Distin, 2005; Stanovich, 2004, 2009). None of these factors are assessed on popular intelligence tests (or their proxies, like the SAT). Intelligence tests do not assess the *propensity* to override responses primed by the autonomous mind or to engage in full cognitive simulation. The crystallized abilities assessed on intelligence tests do not probe for the presence of the specific mindware that is critical for rational thought. And, finally, there are no probes on intelligence tests for the presence of contaminated mindware. Thus we should not be surprised when smart people act foolishly. That we in fact *are* sometimes surprised indicates that we are overvaluing and overconceptualizing the term *intelligence*—we are attributing to it qualities that intelligence tests do not measure. We are missing something important by treating intelligence as if it encompassed all cognitive abilities.

RATIONALITY CAN BE LEARNED AND IRRATIONALITY AMELIORATED

One of the things that we are missing is a focus on the malleability of rationality. This is ironic given that there are at least preliminary indications that rationality may be more malleable than intelligence.

Irrationality caused by mindware gaps is most easily remediable, as it is entirely due to missing strategies and declarative knowledge that can be taught. Overriding the tendencies of the autonomous mind is most often done with learned mindware, and sometimes override fails because of inadequately instantiated mindware. In such a case, inadequately learned mindware is the source of the problem. For example, disjunctive reasoning is the tendency to consider all possible states of the world in deciding among options or choosing a problem solution in a reasoning task. It is a rational thinking strategy with a high degree of generality. People make many suboptimal decisions because of the failure to flesh out all the possible options in a situation, yet the disjunctive mental tendency is not computationally expensive. This is consistent with the finding that there are not strong intelligence-related limitations on the ability to think disjunctively and with evidence indicating that disjunctive reasoning is a rational thinking strategy that can be taught (Adams, 1989; Baron & Brown, 1991; Feehrer & Adams, 1986; Kuhn, 2005; Nickerson, 1988, 2004; Reyna & Farley, 2006; Ritchhart & Perkins, 2005; Swartz & Perkins, 1989).

The tendency to consider alternative hypotheses is, like disjunctive reasoning, strategic mindware of great generality. Also, it can be implemented in very simple ways. Many studies have attempted to teach the importance of thinking of the alternative hypothesis by instructing people in a simple habit. People are given extensive practice at saying to themselves the phrase "think of the opposite" in relevant situations. This strategic mindware does not stress computational capacity and thus is probably easily learnable by many individuals. Several studies have shown that practice at the simple strategy of triggering the thought "think of the opposite" can help to prevent a host of the thinking errors studied in the literature of heuristics and biases, including but not limited to anchoring biases, overconfidence effects, hindsight bias, confirmation bias, and self-serving biases (Arkes, Fault, Guilmette, & Hart, 1988; Koehler, 1994; Koriat, Lichtenstein, & Fischhoff, 1980; Larrick, 2004; Mussweiler, Strack, & Pfeiffer, 2000).

Various aspects of probabilistic thinking represent mindware of great generality and potency. However, as any person who has ever taught a statistics course can attest (the present authors included), some of these insights are counterintuitive and unnatural for people—particularly in their application. There is nevertheless still some evidence that they are indeed teachable—albeit with somewhat more effort and difficulty than strategies such as disjunctive reasoning or considering alternative hypotheses. Aspects of scientific thinking necessary to infer a causal relationship are also definitely teachable (Kuhn, 2005, 2007; Leshowitz, DiCerbo, & Okun, 2002; Nisbett, 1993; Sedlmeier, 1999; Zimmerman, 2007).

Other strategies of great generality may be easier to learn—particularly by those of lower intelligence. For example, psychologist Peter Gollwitzer has discussed an action strategy of extremely wide generality—the use of implementation intentions (Gollwitzer, 1999; Gollwitzer & Schaal, 1998). An implementation intention is formed when the individual marks the cue–action sequence with the conscious, verbal declaration that "When X occurs, I will do Y." The triggering of this cue–action sequence on just a few occasions is enough to establish it in the autonomous mind. Finally, research has shown that an even more minimalist cognitive strategy of forming mental goals (whether or not they have implementation intentions) can be efficacious. For example, people perform better in a task when they are told to form a mental goal ("Set a specific, challenging goal for yourself") for their performance, as opposed to being given the generic motivational instructions ("Do your best"); (Heath, Larrick, & Wu, 1999; Locke & Latham, 1991).

Much of the strategic mindware discussed so far represents learnable strategies in the domain of instrumental rationality (achieving one's goals). Epistemic rationality (having beliefs well calibrated to the world) is often disrupted by contaminated mindware. However, even here, there are teachable macrostrategies that can reduce the probability of acquiring mindware harmful to its host. For example, the principle of falsifiability provides a wonderful inoculation against many kinds of nonfunctional beliefs. It is a tool of immense generality. It is taught in low-level courses on methodology and the philosophy of science, but could be taught much more broadly than this (Stanovich, 2010). Many pseudoscientific beliefs represent the presence of contaminated mindware. The critical thinking skills that help individuals to recognize pseudoscientific belief systems can be taught in high school courses.

If the cognitive miser is easily framed, responds to the most vivid stimulus present, and accepts defaults as given, then the behavior of misers will be shaped by whoever in their world has the power to determine these things (how things are framed, what the most vivid stimulus is, and what the default is). This is clearly problematic, but it suggests that there is another way (other than changing cognition directly) to help people avoid irrational acts. It suggests that a benevolent controller of our environment could help us—could save us from our irrational acts without our having to change basic aspects of our cognition. In short, for certain cognitive problems it might be easier to change the environment than to change people.

For example, in a cross-national study of organ donation rates, Johnson and Goldstein (2006) found that 85.9% of individuals in Sweden had agreed to be organ donors. However, the rate in the United Kingdom was only 17%. The difference in organ donorship between these countries has nothing to do with internal psychological differences between their citizens. The difference is due to a contrast in the public policy about becoming an organ donor in these different countries. In Sweden—like Belgium, France, Poland, and Hungary where agreement to organ donorship is over 95%—the default value on organ donorship is presumed consent. In countries with this public policy, people are assumed to have allowed their organs to be harvested but can opt out by taking an action (usually by getting a notation on their driver's licenses). In contrast, the United States and United Kingdom—like Germany, Denmark, and the Netherlands where agreement to organ donorship is less than 30%—the default value is no donation, with explicit action required to opt *for* organ donation.

In short, the difference between Sweden and the United Kingdom is not in the people. The citizens of both countries are cognitive misers and probably to a roughly equal extent. The great difference is in the form of a particular public policy. As misers, the citizens of both countries are strongly affected by the default heuristic. The option offered as the default is "sticky" in that it is overly influential. A very small change in the donor decision-making environment that hurts no one (since an opt-out procedure is allowed in all countries with presumed consent) could save the lives of thousands of people. The tendencies of the cognitive miser have cost thousands of people their lives. But these tragic consequences are preventable. The best prevention in this case, though, is a change in the environment rather than a change in people, because the former is so much easier to implement.

Thaler and Sunstein (2008) provide another example of environmental fixes by discussing pension participation reforms. The first step comes at the point when employees of most large companies must first choose to enroll. If they do nothing (do not fill out the relevant form) they are not enrolled. Here is where things first go wrong. Many employees do not enroll. New reforms have the employees automatically signed up for the 401(k). They must choose (by filling out a form) to opt out of the system. Such a reform exploits the default bias of the cognitive miser.

An even larger category of problems where people need help from their environments involves their self-control. People overeat, overspend, procrastinate, smoke, and drink too much. Solutions to these problems with self-control are of two forms, corresponding to changes in the individual and changes in the environment. People try to bolster their "will power"—that is, their internal powers of self-control. Alternatively, they try to rearrange their environments so that less exercise of will power (autonomous system override) will be necessary. A common strategy here is to use precommitment devices. People enroll in automatic savings plans so that they will not overspend. They prepackage meals so that they will not overeat. They commit themselves to deadlines so that they will not procrastinate. Precommitments represent our deliberate attempts to restructure our environments so that they will be more conducive to our attempts at self-control.

All of these examples show how simple environmental changes can prevent problems in rational thinking problems. Many more such examples are discussed in Gigerenzer (2002), Stanovich (2009), and Thaler and Sunstein (2008).

RATIONALITY ENCOMPASSES CRITICAL THINKING AND INTELLIGENCE

We can tame intelligence in folk psychology by pointing out that there are legitimate scientific terms as well as folk terms for the other valued parts of cognitive life and that some of these are measurable. This strategy uses to advantage a fact of life that many critics of IQ tests have lamented—that intelligence tests are not going to change any time soon. The tests have the label "intelligence"; thus what they measure will always be dominant in the folk psychology of intelligence. I would argue that it is a mistake to ignore this fact. The tests do not measure rationality, and thus the ability to think rationality will be a subordinate consideration in our schools, in our employment selection devices, and in our society as a whole.

Stanovich (2009) has argued for opening up some space for rationality in the lexicon of the mental and, in doing so, tame the intelligence concept. The term dysrationalia (an analogue of the word *dyslexia*) was defined as the inability to think and behave rationally despite having adequate intelligence (see Stanovich, 1993). Of course, it is easy to recognize that this definition was formulated to contain linguistic and conceptual parallels with the learning disability definitions that stress aptitude–achievement discrepancy. The idea of defining a disability as an aptitude–achievement discrepancy (performance on some domain that is unexpectedly below intelligence) spread widely during the early years of the development of the learning disability concept. Note that the discrepancy idea contains the assumption that all good things should go with high intelligence. When a high score on an IQ test is accompanied by subpar performance in some other domain, this is thought "surprising," and a new disability category is coined to name the surprise.

The strategy in proposing dysrationalia was to prevent intelligence from absorbing the concept of rationality—something that IQ tests do not measure. Restricting the term *intelligence* to what the tests actually measure has the advantage of getting usage in line with the real world of measurement and testing. We have coherent and well-operationalized concepts of rational action and belief formation. We have a coherent and well-operationalized concept of intelligence. No scientific purpose is served by fusing these concepts, because they are very different. To the contrary, scientific progress is made by *differentiating* concepts. Dysrationalia highlights the fact that "all good things" (rationality in this case) do not always go with intelligence. The concept of dysrationalia (and the empirical evidence indicating that the condition is not rare) should help to attenuate our surprise at intelligence–rationality dissociations and to create conceptual space in which we can value abilities at least as important as those presently measured on IQ tests—abilities to form rational beliefs and to take rational action.

The tripartite model of mind presented in this chapter explains why rationality is a more encompassing construct than intelligence. Rationality requires the proper functioning of both the reflective and the algorithmic mind. In contrast, intelligence tests index the computational power of the algorithmic mind. Likewise, the construct of critical thinking is subsumed under the construct of rationality. For example, the processes of critical thinking are often summarized as a set of thinking dispositions that must be developed or inhibited: need for cognition, actively open-minded thinking, belief identification, consideration of future consequences, reflectivity/impulsivity, rational/experiential orientation, need for closure, openness, conscientiousness, etc. These thinking dispositions are the individual difference constructs that capture the functioning of the reflective mind in the tripartite model.

In the context of this model, rationality requires three things: the propensity to override suboptimal responses from the autonomous mind, the algorithmic capacity to inhibit the suboptimal response and to simulate an alternative, and the presence of the mindware that allows the computation of an alternative response. The propensity to override suboptimal responses from the autonomous mind—a property of the reflective mind—captures virtually all of the propensities of critical thinking that have been discussed in the traditional literature on that construct. The algorithmic capacity to inhibit the suboptimal response and to simulate an alternative is captured in standard intelligence tests. Thus, folding critical thinking and intelligence into a generic model of the mind that has rationality as an overarching construct has the considerable advantage of situating, virtually for the first time, the construct of critical thinking within contemporary cognitive science.

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* (pp. 119–141). Mahwah, NJ: Erlbaum.

Adams, M. J. (1989). Thinking skills curricula: Their promise and progress. *Educational Psychologist, 24,* 25–77.

Adams, M. J. (1993). Towards making it happen. *Applied Psychology: An International Review, 42,* 214–218.

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. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin, 110,* 486–498.

Arkes, H., Faust, D., Guilmette, T., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology, 73,* 305–307.

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

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

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.

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

Baron, J. (1993). Why teach thinking?—An essay. *Applied Psychology: An International Review, 42,* 191–214.

Baron, J. (1994). Nonconsequentialist decisions. *Behavioral and Brain Sciences, 17,* 1–42.

Baron, J. (1995). Myside bias in thinking about abortion. *Thinking and Reasoning, 1,* 221–235.

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

Baron, J., & Brown, R. V. (Eds.). (1991). *Teaching decision making to adolescents.* Hillsdale, NJ: Erlbaum.

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.

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

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

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, UK: Cambridge 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* (pp. 29–38). Mahwah, NJ: Erlbaum.

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

Cuban, L. (1984). Policy and research dilemmas in the teaching of reasoning: Unplanned designs. *Review of Educational Research, 54,* 655–681.

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

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). New York: McGraw-Hill.

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

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

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

Dole, J. A., & Sinatra, G. M. (1998). Reconceptualizing change in the cognitive construction of knowledge. *Educational Psychologist, 33,* 109–128.

Donaldson, M. (1978). *Children's minds.* London: Fontana Paperbacks.

Donaldson, M. (1993). *Human minds: An exploration.* New York: Viking Penguin.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51,* 380–417.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. Baron, & R. Sternberg (Eds.). *Teaching thinking skills: Theory and practice* (pp. 9–26). New York: Freeman.

Ennis, R. H. (1996). Critical thinking dispositions: Their nature and assessability. *Informal Logic, 18,* 165–182.

Evans, J. St. B. T. (1972). Interpretation and matching bias in a reasoning task. *Quarterly Journal of Experimental Psychology, 24,* 193–199.

Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences.* Hove, UK: Erlbaum.

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

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

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, UK: Oxford University Press.

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

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

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

Evans, J. St. B. T., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology, 64,* 391–397.

Evans, J. St. B. T., Newstead, S., Allen, J., & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. *European Journal of Cognitive Psychology, 6,* 263–285.

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

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

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, UK: Cambridge University Press.

Feehrer, C. E., & Adams, M. J. (1986). *Odyssey: A curriculum for thinking.* Watertown, MA: Charlesbridge.

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

Foley, R. (1987). *The theory of epistemic rationality.* Cambridge, MA: Harvard University Press.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19,* 25–42.

Garnham, A., & Oakhill, J. V. (2005). Accounting for belief bias in a mental model framework: Comment on Klauer, Musch, and Naumer (2000). *Psychological Review, 112,* 509–518.

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you.* New York: Simon & Schuster.

Gilhooly, K. J., & Murphy, P. (2005). Differentiating insight from non-insight problems. *Thinking and Reasoning, 11,* 279–302.

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.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist, 54,* 493–503.

Gollwitzer, P. M., & Schaal, B. (1998). Metacognition in action: The importance of implementation intentions. *Personality and Social Psychology Review, 2,* 124–136.

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

Hardman, D. (1998). Does reasoning occur on the selection task? A comparison of relevance-based theories. *Thinking and Reasoning, 4,* 353–376.

Harman, G. (1995). Rationality. In E. E. Smith & D. N. Osherson (Eds.), *Thinking* (Vol. 3, pp. 175–211). Cambridge, MA: 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. (2001). *Rational choice in an uncertain world.* Thousand Oaks, CA: Sage.

Heath, C., Larrick, R. P., & Wu, G. (1999). Goals as reference points. *Cognitive Psychology, 38,* 79–109.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26,* 1–23.

Jeffrey, R. C. (1983). *The logic of decision* (2nd ed.). Chicago: University of Chicago Press.

Johnson, E. J., & Goldstein, D. G. (2006). Do defaults save lives? In S. Lichtenstein & P. Slovic (Eds.), *The construction of preference* (pp. 682–688). Cambridge, UK: Cambridge University Press.

Johnson-Laird, P. N. (2006). *How we reason.* Oxford, UK: Oxford 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., & 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. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). Cambridge, UK: Cambridge University Press.

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

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

Kelley, D. (1990). *The art of reasoning.* New York: Norton.

Klaczynski, P. A. (1997). Bias in adolescents' everyday reasoning and its relationship with intellectual ability, personal theories, and self-serving motivation. *Developmental Psychology, 33,* 273–283.

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.

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review, 107*, 852–884.

Koehler, D. J. (1994). Hypothesis generation and confidence in judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 461–469.

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.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107–118.

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

Kuhn, D. (1993). Connecting scientific and informal reasoning. *Merrill-Palmer Quarterly, 38,* 74–103.

Kuhn, D. (2001). How do people know? *Psychological Science, 12,* 1–8.

Kuhn, D. (2005). *Education for thinking.* Cambridge, MA: Harvard University Press.

Kuhn, D. (2007, Feb./March). Jumping to conclusions: Can people be counted on to make sound judgments? *Scientific American Mind,* 44–51.

Kuhn, D., & Udell, W. (2007). Coordinating own and other perspectives in argument. *Thinking & Reasoning, 13,* 90–104.

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

LeBoeuf, R. A., & Shafir, E. (2005). Decision making. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 243–265). New York: Cambridge University Press.

Leshowitz, B., DiCerbo, K. E., & Okun, M. A. (2002). Effects of instruction in methodological reasoning on information evaluation. *Teaching of Psychology, 29,* 5–10.

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

Levesque, H. J. (1986). Making believers out of computers. *Artificial Intelligence, 30,* 81–108.

Levesque, H. J. (1989). Logic and the complexity of reasoning. In R. H. Thomason (Ed.), *Philosophical logic and artificial intelligence* (pp. 73–107). Kluwer.

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.

Lipman, M. (1991). *Thinking in education.* Cambridge, UK: Cambridge University Press.

Locke, E. A., & Latham, G. P. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes, 50,* 212–247.

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, UK: Psychology Press.

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

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Matthews, G., Zeidner, M., & Roberts, R. D. (2002). *Emotional intelligence: Science & 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.

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.

Moshman, D. (2004). From inference to reasoning: The construction of rationality. *Thinking and Reasoning, 10,* 221–239.

Moshman, D. (2005). *Adolescent psychological development: Rationality, morality, and identity* (2nd ed.). Mahwah, NJ: Erlbaum.

Moshman, D. (2010). The development of rationality. In H. Siegel (Ed.), *Oxford Handbook of Philosophy of Education* (pp. 145–161). Oxford, UK: Oxford University Press.

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, UK: Cambridge University Press.

Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin, 9,* 1142–1150.

Neimark, E. (1987). *Adventures in thinking.* San Diego, CA: Harcourt Brace Jovanovich.

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.

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

Nickerson, R. (1987). Why teach thinking? In J. Baron & R. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 27–40). New York: Freeman.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2,* 175–220.

Nickerson, R. S. (2004). *Cognition and chance: The psychology of probabilistic reasoning.* Mahwah, NJ: Erlbaum.

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

Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Midwest Publications.

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 Publishing.

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

Paul, R. W. (1984). Critical thinking: Fundamental to education for a free society in North America: A new theory of knowledge, learning, and literacy. *Educational Leadership, 42,* 4–14.

Paul, R. W. (1987). Critical thinking and the critical person. In D. N. Perkins, J. Lockhead & J. Bishop (Eds.), *Thinking: The second international conference* (pp. 373–403). Hillsdale, NJ: Erlbaum.

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

Perkins, D. N. (2002). The engine of folly. In R. J. Sternberg (Ed.), *Why smart people can be so stupid* (pp. 64–85). New Haven, CT: Yale University Press.

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

Piaget, J. (1926). *The language and thought of the child.* London: Routledge.

Reyna, V. F. (2004). How people make decisions that involve risk. *Current Directions in Psychological Science, 13,* 60–66.

Reyna, V. F., & Farley, F. (2006). Risk and rationality in adolescent decision making. *Psychological Science in the Public Interest, 7,* 1–44.

Ritchhart, R., & Perkins, D. N. (2005). Learning to think: The challenges of teaching thinking. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 775–802). New York: Cambridge University 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.

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, UK: Oxford University Press.

Savage, L. J. (1954). *The foundations of statistics.* New York: Wiley.

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

Sedlmeier, P. (1999). *Improving statistical reasoning: Theoretical models and practical implications.* Mahwah, NJ: Erlbaum.

Shafir, E. (1994). Uncertainty and the difficulty of thinking through disjunctions. *Cognition, 50,* 403–430.

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

Siegel, H. (1988). *Educating reason.* New York: Routledge.

Siegel, H. (1997). *Rationality redeemed? Further dialogues on an educational ideal.* New York: Routledge.

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.

Stanovich, K. E. (1993). Dysrationalia: A new specific learning disability. *Journal of Learning Disabilities, 26,* 501–515.

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

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: University of Chicago Press.

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

Stanovich, K. E. (2010). *How to think straight about psychology* (9th ed.). Boston: Allyn & Bacon.

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 & 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. (2007). Natural myside bias is independent of cognitive ability. *Thinking & 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 & 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.

Sternberg, R. J. (1997). *Thinking styles.* Cambridge, UK: Cambridge University Press.

Sternberg, R. J. (2001). Why schools should teach for wisdom: The balance theory of wisdom in educational settings. *Educational Psychologist, 36,* 227–245.

Sternberg, R. J. (2002). Smart people are not stupid, but they sure can be foolish: The imbalance theory of foolishness. In R. J. Sternberg (Ed.), *Why smart people can be so stupid* (pp. 232–242). New Haven, CT: Yale University Press.

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

Sternberg, R. J. (2005). Foolishness. In R. J. Sternberg & J. Jordan (Eds.), *A handbook of wisdom: Psychological perspectives* (pp. 331–352). New York: Cambridge University Press.

Sternberg, R. J. (2006). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical, and creative skills. *Intelligence, 34,* 321–350.

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

Sternberg, R. J., & Jordan, J. (Eds.). (2005). *A handbook of wisdom: Psychological perspectives.* New York: Cambridge University Press.

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.

Swartz, R. J., & Perkins, D. N. (1989). *Teaching thinking: Issues & approaches.* Pacific Grove, CA: Midwest Publications.

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

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* New Haven, CT: Yale University Press.

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.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211,* 453–458.

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

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

Tweney, R. D., & Yachanin, S. (1985). Can scientists rationally assess conditional inferences? *Social Studies of Science, 15,* 155–173.

von Neumann, J., & Morgenstern, O. (1944). *The theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Voss, J., Perkins, D., & Segal J. (Eds.). (1991). *Informal reasoning and education.* Hillsdale, NJ: Erlbaum.

Wade, C., & Tavris, C. (1993). *Critical and creative thinking.* New York: HarperCollins.

Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Harmondsworth, UK: Penguin.

West, R. F., & Stanovich, K. E. (2003). Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory & Cognition, 31,* 243–251.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117–142.

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.

Zechmeister, E. B., & Johnson, J. (1992). *Critical thinking: A functional approach.* Pacific Grove, CA: Brooks/Cole.

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.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27,* 172–223.