Rationality, Intelligence, and Levels of Analysis in Cognitive Science: Is Dysrationalia Possible?

Keith E. Stanovich

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.

Rationality, Intelligence, and Levels of Analysis in Cognitive Science: Is Dysrationalia Possible?

In a 1994 article in the journal Cognition, Eldar Shafir describes a very straightforward rule from decision theory. The rule was termed the sure-thing principle by Savage (1954) and it says the following. Imagine you are choosing between two possible outcomes, A and B, and event X is an event that may or may not occur in the future. If you prefer prospect A to prospect B if X happens and also you prefer prospect A to prospect B if X does not happen, then you definitely prefer A to B and that preference is in no way changed by knowledge of event X so you should prefer A to B whether you know anything about event X or not. Shafir (1994) calls the sure-thing principle "one of simplest and least controversial principles of rational behavior" (p. 404). Indeed, it is so simple and obvious that it seems hardly seems worth stating. Yet Shafir (1994), in his article, reviews a host of studies that have demonstrated that people do indeed violate the sure-thing principle.

For example, Tversky and Shafir (1992) created a scenario where subjects were asked to imagine that they were at the end of the term, tired and run down, and awaiting the grade in a course that they might fail and be forced to retake. They were to imagine that they had just been given the opportunity to purchase an extremely attractive vacation package to Hawaii at a very low price. More than half of a group of students who were informed that they had passed the exam chose to buy the vacation package and an even larger proportion of a group who had been told that they had failed the exam chose to buy the vacation package. However, only one third of a group who did not know whether they passed or failed the exam chose to purchase the vacation.

What these results collectively mean is that, by inference, at least some subjects were saying, essentially "I'll go if I pass and I'll go if I fail, but I won't go if I don't know whether I passed or failed."

Shafir (1994) describes a host of decision situations where this outcome obtains. Subjects prefer A to B when event X obtains, prefer A to B when X does not obtain, but prefer B to A when uncertain about the outcome X--a clear violation of the sure-thing principle. These violations are not limited to toy problems or laboratory situations. Shafir (1994) provides some real-life examples, one involving the stock market just prior to Bush/Dukakis election of 1988. Market analysts were near unanimous in their opinion that Wall Street preferred Bush to Dukakis. Yet subsequent to Bush's election, stock and bond prices declined and the dollar plunged to its lowest level in ten months. Of course, analysts agreed that the outcome would have been worse had Dukakis been elected. Yet if the market was going to go down subsequent to the election of Bush, and going to go down even further subsequent to the election of Dukakis, then why didn't it go down before the election due to the absolute certainty that whatever happened (Bush or Dukakis) the outcome was bad for the market! The market seems to have violated the sure-thing principle.

The sure-thing principle is not the only rule of rational thinking that humans have been shown to violate. A substantial research literature--one comprising literally hundreds of empirical studies conducted over nearly four decades--has firmly established that people's responses often deviate from the performance considered normative on many reasoning tasks. For example, people assess probabilities incorrectly, they display confirmation bias, they test hypotheses inefficiently, they violate the axioms of utility theory, they do not properly calibrate degrees of belief, they overproject their own opinions onto others, they display illogical framing effects, they uneconomically honor sunk costs, they allow prior knowledge to become implicated in deductive reasoning, and they display numerous other information processing biases (for summaries of the large literature, see Arkes, 1991; Baron, 1994, 1998; Dawes, 1998; Evans, 1989; Evans & Over, 1996; Kahneman, Slovic, & Tversky, 1982; Nickerson, 1998; Osherson, 1995; Piattelli-Palmarini, 1994; Plous, 1993; Shafir & Tversky, 1995; Stanovich, 1999; Tversky, 1996).

The reader need not be familiar with all of these principles of rational thinking. It is sufficient to appreciate that many of them are as fundamental as the sure-thing principle just discussed. It is also important to point out that these reasoning errors do cash out in real-life behaviors that are decidedly suboptimal and unpleasant for those displaying these processing biases. Because of the failure to follow the normative rules of rational thought--because of the processing biases listed above--physicians choose less effective medical treatments (McNeil, Pauker, Sox, & Tversky, 1982; Redelmeier & Tversky, 1990, 1992; Sutherland, 1992); people fail to accurately assess risks in their environment (Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978; Margolis, 1996; Yates, 1992); information is misused in legal proceedings (Saks & Kidd, 1980–1981); millions of dollars are spent on unneeded projects by government and private industry (Arkes & Ayton, 1999; Dawes, 1988, pp. 23–24); parents fail to vaccinate their children (Baron, 1998); unnecessary surgery is performed (Dawes, 1988, pp. 73–75); animals are hunted to extinction (Baron, 1998; Dawkins, 1998); billions of dollars are wasted on quack medical remedies (Gilovich, 1991); and costly financial misjudgments are made (Belsky, 1995; Belsky & Gilovich, 1999; Fridson, 1993; Thaler, 1992; Tversky, 1996; Willis, 1990).

Many of these examples concern what philosophers call pragmatic, or practical rationality-- how well a person maximizes the satisfaction of their desires, given their beliefs (Audi, 1993; Harman, 1995; Nathanson, 1994). This is often contrasted with epistemic rationality which is concerned with the consistency of a person's network of beliefs and how well it represents the external world (the so-called theoretical rationality of philosophy: Audi, 1993; Foley, 1987; Harman, 1995).

How Are We Understand Smart People Doing Dumb Things?

The findings from the reasoning and decision making literature and the many real-world examples of the consequences of irrational thinking (e.g., Belsky & Gilovich, 1999; Gilovich, 1991; Piattelli-Palmarini, 1994; Shermer, 1997; Sutherland, 1992; Thaler, 1992) create a seeming paradox. The physicians using ineffective procedures, the financial analysts making costly misjudgments, the retired professionals managing their money poorly--none of these are unintelligent people. The experimental literature is even more perplexing. Over 90% of the subjects in the studies in the literature are university students--some from the most selective institutions of higher learning in the world (Tversky & Shafir's subjects are from Stanford). Yet these are the very people who have provided the data that indicate that a substantial proportion of people can sometimes violates some of the most basic strictures of rational thought such as transitivity or the sure-thing principle. It appears that an awful lot of pretty smart people are doing some incredibly dumb things. How are we to understand this seeming contradiction?

The first step in understanding the seeming paradox is to realize that the question "How can so many smart people be doing so many dumb things?" is phrased in the language of folk psychology. The issue of how to interpret folk psychology is a topic of immense interest in cognitive science at present, and it is the subject of much controversy (Christensen

& Turner, 1993; Churchland & Churchland, 1998; Davies & Stone, 1995; Greenwood, 1991; Stich, 1996). Positions vary from those who think that most folk psychology needs to eliminated from the terminology of scientific psychology to those who think that folk psychology should be the very foundation of a scientific psychology. My concern here is not with these classic issues but with how concepts in cognitive science can be used to sharpen up folk usage in ways that serve to dissipate seeming paradoxes[1]. I propose to do just this with the "smart but dumb" phrase. In this chapter, I identify the folk term "smart" with the psychology concept of intelligence (defined as an amalgamation of cognitive capacities). The acts that spawn the folk term "dumb" I identify with violations of rationality as that term is conceptualized within cognitive science, philosophy, and decision science (Baron, 1993a; Harman, 1995; Jeffrey, 1983; Kleindorfer, Kunreuther, & Schoemaker, 1993; Nathanson, 1994; Nozick, 1993). This mapping does not immediately solve the problem because there are several different ways of parsing the concepts intelligence and rationality--especially within psychology. Thus, I present one such partitioning that I think is useful in contextualizing the "smart but dumb" phenomenon and dissolving its seemingly paradoxical status. The partitioning that I prefer relies heavily on distinguishing levels of analysis in cognitive theory.

<u>Levels of Analysis in Cognitive Science</u>

Levels of analysis in cognitive theory have been discussed by numerous theorists (Anderson, 1990, 1991; Dennett, 1978, 1987; Horgan & Tienson, 1993; Levelt, 1995; Marr, 1982; Newell, 1982, 1990; Oaksford & Chater, 1995; Pylyshyn, 1984; Sterelny, 1990). For example, Anderson (1990) defines four levels of theorizing in cognitive science: a biological level that is inaccessible to cognitive theorizing; an implementation level which is basically a comprehensible shorthand approximation to the biological; an algorithmic level concerned with the computational processes necessary to carry out a task; and the rational level. The latter level provides a specification of the goals of the system's computations (what the system is attempting to compute and why) and can be used to suggest constraints on the operation of the algorithmic level. The rational level of analysis is concerned with the goals of the system, beliefs relevant to those goals, and the choice of action that is rational given the system's goals and beliefs (Bratman, Israel, & Pollack, 1991; Dennett, 1987; Newell, 1982, 1990; Pollock, 1995).

Many similar taxonomies exist in the literature (Sterelny, 1990, p. 46, warns of the "bewildering variety of terms" used to describe these levels of analysis). Indeed, Anderson's (1990) draws heavily on the work of Marr (1982) and Newell (1982). Table 1 presents the alternative, but similar, schemes of Anderson (1990), Marr (1982), Newell (1982), Dennett (1987), and a compromise scheme that I used in a 1999 volume (Stanovich, 1999) and that will be used in this chapter. The first level of analysis is termed the biological level in my taxonomy because I will be largely concerned with human information processing rather than computational devices in general. My scheme follows Marr (1982) and Dennett (1987) in collapsing Anderson's algorithmic and implementation levels into one because for the purposes of the present discussion the distinction between these two levels is not important. This second level is termed algorithmic--a term that is relatively uncontroversial.

Insert Table 1 about here

In contrast, the proper term for the third level is variable and controversial. Borrowing from Dennett (1987), I have termed this level of analysis the intentional level for the following reasons. First, Anderson

(1990) has argued that Marr's (1982) terminology is confusing and inapt because "his level of computational theory is not really about computation but rather about the goals of the computation. His basic point is that one should state these goals and understand their implications before one worries about their computation, which is really the concern of the lower levels of his theory" (p. 6). Dennett (1987) reiterates this critique of Marr's terminology by noting that "the highest level, which he misleadingly calls computational, is in fact not at all concerned with computational processes but strictly (and more abstractly) with the question of what function the system in question is serving" (pp. 74-75). The term chosen by Newell (1982)--the knowledge level--is equally inapt in not signaling that this level is concerned with action selection based on expected goal attainment in light of current beliefs. Instead, I have adapted Dennett's terminology and referred to this level as the intentional level of analysis. Although Sterelny (1990, p. 45) argues that this level of analysis is not necessarily tied to an intentional psychology, like Dennett, I do want to conjoin the two--so in the present case, the term is apt.

Thinking Dispositions, Cognitive Capacities, and Levels of Analysis

In many areas of psychology there is increasing attention being paid to behavioral/cognitive concepts that reside at the borderline of cognitive psychology and personality (Ackerman & Heggestad, 1997; Goff & Ackerman, 1992; Haslam & Baron, 1994; Keating, 1990; Nickerson, 1988; Perkins, 1995; Perkins, Jay, & Tishman, 1993; Rolfhus & Ackerman, 1999; Siegel, 1993; Stanovich & West, 1997; Sternberg, 1997b; Sternberg & Ruzgis, 1994; Swartz & Perkins, 1989). Moshman (1994), for instance, reminds us of the importance of "considerations of will and disposition [because they] lie at the interface of cognition with affect, motivation, social relations, and cultural context" (p.143), and Sternberg (1988) likewise notes that "intellectual styles represent an important link between intelligence and personality, because they probably represent, in part, a way in which personality is manifested in intelligent thought and action" (p. 218). Terminology surrounding such notions is remarkably varied. The term thinking dispositions will be used in this chapter (see Baron, 1988; Ennis, 1987; Perkins, 1995; Stanovich & West, 1997), although other theorists--in dealing with similar concepts--prefer terms such as intellectual style (Sternberg, 1988, 1989), cognitive emotions (Scheffler, 1991), habits of mind (Keating, 1990), inferential propensities (Kitcher, 1993, pp. 65-72), epistemic motivations (Kruglanski, 1990), constructive metareasoning (Moshman, 1994), styles of epistemic regulation (Sá, West, & Stanovich, 1999; Stanovich, 1999); cognitive styles (Messick, 1984, 1994), and thinking styles (Sternberg, 1997b). Despite this diversity of terminology, most authors use such terms similarly--to refer to relatively stable psychological mechanisms and strategies that tend to generate characteristic behavioral tendencies and tactics (see Buss, 1991).

In this chapter, it is proposed that thinking dispositions should be distinguished from cognitive capacities because the two constructs are at different levels of analysis in cognitive theory and do separate explanatory work. This distinction motivates interest in a consistent empirical finding in the literature--that thinking dispositions can predict performance on reasoning and rational thinking tasks even after individual differences in measures of general cognitive ability have been partialled out.

The distinction between cognitive capacities and thinking dispositions has been drawn by many theorists (e.g., Baron, 1985, 1988, 1993b; Ennis, 1987; Moshman, 1994; Norris, 1992; Perkins et al., 1993; Schrag, 1988). For example, in Baron's (1985, 1988) conceptualization, capacities refer to the types of cognitive processes studied by information processing researchers seeking the underlying cognitive basis of performance on IQ tests. Perceptual speed, discrimination accuracy, working memory capacity, and the efficiency of the retrieval of information stored in long-term memory are examples of cognitive capacities that underlie traditional psychometric intelligence and that have been extensively investigated (Ackerman, Kyllonen, & Richards, 1999; Carpenter et al., 1990; Deary & Stough, 1996; Engle, Tuholski, Laughlin, & Conway, 1999; Fry & Hale, 1996; Hunt, 1978, 1987; Lohman, 1989; Sternberg, 1982; Vernon, 1991, 1993). These cognitive capacities are what Baltes (1987) terms the "mechanics of intelligence." Psychometric g provides an overall index of the cognitive efficiency of a wide variety of such mechanisms in a given individual (Carroll, 1993, 1997). According to Baron's (1985) conception, cognitive capacities cannot be improved in the short-term by admonition or instruction. They are, nevertheless, affected by long-term practice.

Thinking dispositions, in contrast, are better viewed as cognitive styles which are more malleable: "Although you cannot improve working memory by instruction, you can tell someone to spend more time on problems before she gives up, and if she is so inclined, she can do what you say" (Baron, 1985, p. 15). Rational thinking dispositions are those that relate to the adequacy of belief formation and decision making, things like "the disposition to weigh new evidence against a favored belief heavily (or lightly), the disposition to spend a great deal of time (or very little) on a problem before giving up, or the disposition to weigh heavily the opinions of others in forming one's own" (Baron, 1985, p. 15).

By and large, psychometric instruments such as IQ tests have tapped cognitive capacities almost exclusively and have ignored cognitive styles and thinking dispositions (Baron, 1985, 1988; Stanovich, 1994; Sternberg, 1997b). Importantly, Baron (1988) argues that, in ignoring dispositions, the

IQ concept "has distorted our understanding of thinking. It has encouraged us to believe that the only general determinants of good thinking are capacities, and this attitude has led to the neglect of general dispositions" (p. 122; see also Sternberg 1997b). It will be argued here that the study of thinking dispositions balances this tendency by directing attention to the possibility of systematically suboptimal systems at the intentional level of analysis.

Recall that each level of analysis in cognitive theory frames a somewhat different issue. At the algorithmic level the key issue is one of computational efficiency, and at the biological level the paramount issue is whether the physical mechanism has the potential to instantiate certain complex algorithms. In contrast, it is at the intentional level that issues of rationality arise. Using this taxonomy, it is proposed here that omnibus measures of cognitive capacities such as intelligence tests are best understood as indexing individual differences in the efficiency of processing at the algorithmic level. In contrast, thinking dispositions as traditionally studied in psychology (e.g., Cacioppo, Petty, Feinstein, & Jarvis, 1996; Kardash & Scholes, 1996; Kruglanski & Webster, 1996; Klaczynski, Gordon, & Fauth, 1997; Schommer, 1990, 1993, 1994; Stanovich & West, 1997; Sternberg, 1997b) index individual differences at the intentional level of analysis. They are telling us about the individual's goals and epistemic values (Sá et al., 1999)--and they are indexing broad tendencies of pragmatic and epistemic self-regulation. For example, in his model of mind as a control system, Sloman (1993) views desires as control states that can produce behavior either directly or through a complex control hierarchy by changing intermediate desire-states. He views dispositions (high-level attitudes, ideals, and personality traits) as long-term desire states that "work through a control hierarchy, for instance, by changing other desire-like states rather than triggering behaviour" (p. 85).

Thus, thinking dispositions are reflective of intentional-level psychological structure. It has been the goal of our research program to determine whether such features of intentional-level psychology can serve as explanatory mechanisms in accounts of discrepancies between normative and descriptive models of behavior (Stanovich, 1999). If thinking dispositions correlate with individual differences in the normative/descriptive gap then this will be prima facie evidence that the gap is caused by actual differences in intentional psychology. However, any such association might well arise because the variation in thinking dispositions is co-extensive with differences in computational capacity. Thus, it will be important to examine whether intentional-level cognitive dispositions can explain unique variance--variance independent of cognitive capacity. This has been one of the major analytic tests that we have used when examining individual differences across a variety of rational thinking tasks in the heuristics and biases literature (Stanovich, 1999). In short, we have been searching for systematic differences in intentional-level psychology that are not explainable by variation in algorithmic capacity.

Thinking Dispositions as Predictors of Rational Thought

Discussions of critical thinking in the educational and psychological literature consistently point to the importance of the ability to evaluate arguments and evidence in a way that is not contaminated by one’s prior beliefs. For example, Norris and Ennis (1989) list as one characteristic of critical thinkers the tendency to “reason from starting points with which they disagree without letting the disagreement interfere with their reasoning” (p. 12). Similarly, Nickerson (1987) and many other theorists (e.g., Lipman, 1991; Paul, 1984, 1987; Perkins, 1995; Perkins et al., 1993; Swartz & Perkins, 1989) stress 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” (Nickerson, 1987, p. 30). The growing literature on informal or practical reasoning likewise emphasizes the importance of detaching one’s own beliefs from the process of argument evaluation (Baron, 1991, 1995; Brenner, Koehler, & Tversky, 1996; Kardash & Scholes, 1996; Klaczynski & Gordon, 1996; Klaczynski et al., 1997; Kuhn, 1991, 1993; Perkins, 1985; Stanovich & West, 1997; Voss, Perkins, & Segal, 1991).

In light of the emphasis in the critical thinking literature on the importance of evaluating arguments independently of prior belief, it is noteworthy that there are increasing indications in the research literature that individual differences in this skill can be predicted by thinking dispositions even after differences in general cognitive ability have been partialled out. For example, Schommer (1990) found that a measure of the disposition to believe in certain knowledge predicted the tendency to draw one-sided conclusions from ambiguous evidence even after verbal ability was controlled. Kardash and Scholes (1996) found that the tendency to properly draw inconclusive inferences from mixed evidence was related to belief in certain knowledge and to a measure of need for cognition (Cacioppo, et al., 1996). Furthermore, these relationships were not mediated by verbal ability because a vocabulary measure was essentially unrelated to evidence evaluation. Likewise, Klaczynski (1997; see also Klaczynski & Gordon, 1996; Klaczynski et al., 1997) found that the degree to which adolescents criticized belief-inconsistent evidence more than belief-consistent evidence was unrelated to cognitive ability (see also, Perkins, Farady, & Bushey, 1991).

Results from our own studies have converged with those of Schommer

(1990) and Kardash and Scholes (1996) in indicating that thinking dispositions can predict argument evaluation skill once cognitive ability is partialled out. We have developed an argument evaluation task in which we derive an index of the degree to which argument evaluation is associated with argument quality independent of prior belief (see Stanovich & West, 1997; Sá et al., 1999). Our methodology involves assessing, on a separate instrument, the participant's prior beliefs about a series of controversial propositions. On an argument evaluation measure, administered at a later time, the participants evaluate the quality of arguments related to the same propositions. The arguments have an operationally determined objective quality that varies from item to item. Our analytic strategy is to regress each subject's evaluations of the argument simultaneously on the objective measure of argument quality and on the strength of the belief he/she had about the propositions prior to reading the argument. The standardized beta weight for argument quality then becomes an index of that subject's reliance on the quality of the arguments independent of the subject's beliefs on the issues in question. The magnitude of the former statistic becomes an index of argument-driven, or data-driven processing (to use Norman's [1976] term).

Our methodology is different from the traditional logic used in critical thinking tests, and it is a more sensitive one for measuring individual differences (see Stanovich, 1999, for a discussion). For example, standardized critical thinking tests often simply try to balance opinions across items by utilizing a variety of issues and relying on chance to ensure that prior belief and strength of the argument are relatively balanced from respondent to respondent (Watson & Glaser, 1980). In contrast, we actually <u>measured</u> the prior opinion and took it into account in the analysis (for related techniques, see Klaczynski & Gordon, 1996; Kuhn, 1991, 1993;

Kuhn, Amsel, & O'Loughlin, 1988; Slusher & Anderson, 1996). The technique allowed us to examine thought processes in areas of "hot" cognition where biases are most likely to operate (Babad & Katz, 1991; Klaczynski & Narasimham, 1998; Kunda, 1990, 1999; Pyszczynski & Greenberg, 1987).

We have consistently found (see Stanovich & West, 1997; Sá et al., 1999) that, even after controlling for cognitive ability, individual differences on our index of argument-driven processing can be predicted by measures of dogmatism and absolutism (Rokeach, 1960), categorical thinking (Epstein & Meier, 1989), openness (Costa & McCrae, 1992), flexible thinking (Stanovich & West, 1997), belief identification (Sá et al., 1999), counterfactual thinking, superstitious thinking (Stanovich, 1989; Tobacyk & Milford, 1983), and actively open-minded thinking as conceptualized by Baron (1985, 1988, 1993b; see also, Facione, 1992; Norris & Ennis, 1989; Perkins et al., 1993). These findings converge with those of Schommer (1990) and Kardash and Scholes (1996) in supporting a conceptualization of human cognition that emphasizes the potential separability of cognitive capacities and thinking styles/dispositions as predictors of reasoning skill (e.g., Baron, 1985, 1988; Ennis, 1987; Kitchener & Fischer, 1990; Klaczynski, et al., 1997; Norris, 1992; Schrag, 1988; Siegel, 1993; Sternberg, 1997b).

Such a separation in psychological constructs makes sense if indeed they do map on to different levels of analysis in cognitive theory. I proposed earlier that variation in cognitive ability refers to individual differences in the efficiency of processing at the algorithmic level. In contrast, thinking dispositions index individual differences at the intentional level. They are telling us about the individual's goals and epistemic values (King & Kitchener, 1994; Kitcher, 1993; Kruglanski & Webster, 1996; Pintrich, Marx, & Boyle, 1993; Schommer, 1990, 1993, 1994; Stanovich, 1999). For example, consider an individual who scores high on our measures of actively open-minded thinking (see Stanovich & West, 1997) and low on measures of dogmatism and absolutism--a person who agrees with statements such as "People should always take into consideration evidence that goes against their beliefs" and who disagrees with statements such as "No one can talk me out of something I know is right." Such a response pattern is indicating that this person values belief change in order to get closer to the truth. This individual is signaling that they value being an accurate belief forming system more than they value holding on to the beliefs they currently have (see Cederblom, 1989 for an insightful discussion of this distinction and our scale based on this notion in Sá et al., 1999).

In contrast, consider a person scoring low on actively open-minded thinking measures and high on measures of absolutism and categorical thinking--a person who disagrees with statements such as "A person should always consider new possibilities" and who agrees with statements such as "There are a number of people I have come to hate because of the things they stand for." Such a response pattern is indicating that retaining current beliefs is an important goal for this person. This individual is signaling that they value highly the beliefs they currently have and that they put a very small premium on mechanisms that might improve belief accuracy (but that involve belief change).

In short, thinking dispositions of the type studied by Schommer (1990, 1993, 1994), Kardash and Scholes (1996), and Stanovich and West (1997) provide information about epistemic goals at the rational level of analysis. Within such a conceptualization, we can perhaps better understand why such thinking dispositions predict additional variance in argument evaluation even after cognitive ability is partialled out. This result may be indicating that to understand variation in reasoning in such a task we need to examine more than just differences at the algorithmic level (computational capacity)--we must know something about the epistemic goals of the reasoners.

Thus, performance on tasks requiring reasoning about previously held beliefs, while certainly somewhat dependent upon the cognitive capacity of the subject, also depends on the balance of epistemic goals held by the reasoners. The instructions for many tasks which require reasoning in the face of belief bias (Baron, 1995; Evans, Newstead, Allen, & Pollard, 1994; Oakhill, Johnson-Laird, & Garnham, 1989; Stanovich & West, 1997) dictate that prior belief be totally discounted in evaluating the argument. But individuals may differ in their willingness and/or ability to adapt to such instructions. Some individuals may put a low priority on allocating computational capacity to evaluate the argument. Instead, for them, capacity is engaged to assess whether the conclusion is compatible with prior beliefs (Evans, Barston, & Pollard, 1983; Evans et al., 1994). Other individuals--of equal cognitive ability--may marshal their cognitive resources to decouple (see Navon, 1989a, 1989b) argument evaluation from their prior beliefs as the instructions demand. These individuals may easily engage in such a processing strategy because it does not conflict with their epistemic goals. Many problems in practical reasoning may have a similar logic (Baron, 1991, 1995; Foley, 1991; Klaczynski & Gordon, 1996; Kuhn, 1991; Perkins et al., 1991; Schoenfeld, 1983). Such problems--although they obviously stress algorithmic capacity to varying degrees--might also differ greatly in how they engage people's goal structure.

Thus, to fully understand variation in evidence evaluation performance, we might need to consider variation at the rational level as well as at the algorithmic level of cognitive analysis. Indeed, this may be true for other measures of rational and critical thought as well. In fact, we have linked various measures of thinking dispositions to statistical reasoning tasks of various types (Stanovich, 1999; Stanovich & West, 1998, 2000). For example, Nisbett and Ross (1980) have reviewed the evidence on the tendency of human judgment to be overly influenced by vivid but unrepresentative personal and testimonial evidence and to be under-influenced by more representative and diagnostic statistical evidence. Studying the variation in this response tendency is important because Griffin and Tversky (1992) argue that "the tendency to prefer an individual or 'inside' view rather than a statistical or 'outside' view represents one of the major departures of intuitive judgment from normative theory" (pp. 431-432). The quintessential problem (see Fong, Krantz, & Nisbett, 1986) involves choosing between contradictory car purchase recommendations--one from a large-sample survey of car buyers and the other the heartfelt and emotional testimony of a single friend. Fong et al. (1986) and Jepson, Krantz, & Nisbett (1983) have studied a variety of such problems and we have examined a number of them in our own research. We have consistently found that, even though these problems are presented to participants as having no right or wrong answers, dispositions toward actively open-minded thinking (Baron, 1993b) are consistently associated with reliance on the statistical evidence rather than the testimonial evidence. Furthermore, this association remains even after cognitive ability has been controlled.

We have examined a variety of other critical and rational thinking tasks and have consistently found the same pattern. For example, we have examined the phenomenon of outcome bias in decision evaluation (Baron & Hershey, 1988)--the tendency to rate decision quality according to the outcome of the decision even when the outcome provides no cues to the information available to the decision maker. We again found that the ability to avoid outcome bias was associated with dispositions toward actively open-minded thinking and that this tendency was not due solely to differences in cognitive ability. Similar results were found for a variety of other hypothesis testing and reasoning tasks (Stanovich, 1999; Stanovich & West, 1998, 2000).

Throughout several of our studies normative responding on a variety of problems from the heuristics and biases literature (see Arkes & Hammond, 1986; Kahneman & Tversky, 2000; Kahneman, Slovic, & Tversky, 1982) was moderately correlated with cognitive ability. Nevertheless, these algorithmic-level limitations were far from absolute. The magnitude of the associations with cognitive ability left much room for the possibility that the remaining reliable variance might indicate that there are systematic irrationalities in intentional-level psychology. It was rarely the case that once capacity limitations had been controlled, that the remaining variations from normative responding were unpredictable (which would have indicated that the residual variance consisted largely of performance errors). In several studies, we have shown that there was significant covariance among the scores from a variety of tasks in the heuristics and biases literature after they had been residualized on measures of cognitive ability (Stanovich, 1999). The residual variance (after partialling cognitive ability) was also systematically associated with questionnaire responses that were conceptualized as intentional-level styles relating to epistemic regulation (Sá et al. 1999; Stanovich, 1999; Stanovich & West, 1997, 1998, 2000). Both of these findings are indications that the residual variance is systematic. They falsify models that attempt to explain the normative/descriptive gap entirely in terms of computational limitations and random performance errors. Instead, the findings support the notion that the normative/descriptive discrepancies that remain after computational limitations have been accounted for reflect a systematically suboptimal intentional-level psychology.

The Rationality/Intelligence Demarcation: Dissolving the Smart But Dumb Paradox

The empirical work just summarized illustrates why I think that the distinction between cognitive capacities and thinking dispositions is useful to psychological theory. I further propose that it might clarify folk usage like the "smart but dumb" phrasing if the concept of intelligence be restricted to the domain of individual differences in cognitive capacities--in short, that the term be restricted to discussions of computational capacity at the algorithmic level of analysis. In contrast, the term rationality is used here to refer to styles of epistemic and response regulation at the intentional level of analysis.

Using this terminology, we see that the results summarized above can be taken to indicate that while the algorithmic level constrains the intentional level--as it is standard to assume in cognitive science (Cherniak, 1986; Goldman, 1978; Harman, 1995; Oaksford & Chater, 1993, 1995)--from an individual differences perspective, the correlation between individual differences at the two levels is less than unity. Thus, dissociations between the intentional-level individual differences and algorithmic-level individual differences are indeed possible. Rationality can dissociate from intelligence on this view.

A little mapping of of folk psychological terms now resolves the paradox with which we opened this chapter. In the vernacular, we often say "what a dumb thing to do" when irrational thinking has led to a maladaptive behavioral act--an act best analyzed by positing suboptimal action regulation at the intentional level of analysis. For example, Baron

(1985) notes that "When we disapprovingly call a person 'stupid' because of some action, for example, a political leader, we do not often mean that the action was done too slowly, or that it would not have been done if the doer had a larger working memory capacity....When we call someone stupid, we are really saying he is irrational, not that he is retarded" (p. 235).

The problem here is that for many, the antonym for dumb and stupid is often "smart" and this, to most people, often connotes intelligence--here viewed as an algorithmic-level concept having to do with cognitive capacity. If the folk psychological view shades the connotation of "dumb" a little more towards rationality than toward intelligence, and the connotation of smart a little less toward rationality and a little more toward intelligence, then there is no paradox at all. There is nothing strange in smart people acting dumb because people can have considerable algorithmic capacity yet still display irrational behavior and thought because of systematic suboptimalities in their intentional-level psychologies--in the systems that regulate epistemic functioning and action determination (see Stanovich, 1999). This is one way to view what the heuristics and biases literature has been demonstrating now for over 30 years (e.g., Arkes, 1991; Dawes, 1998; Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 2000; Nickerson, 1998; Piattelli-Palmarini, 1994; Shafir & Tversky, 1995). In short, if the folk usage is parsed in this manner, then there is a 30-year research history of demonstrations of "smart people acting dumb".

Dysrationalia: Demarcating Intelligence and Rationality

In previous publications (Stanovich, 1993, 1994) I have tried to draw attention to intelligence/rationality distinction by proposing a new discrepancy-based disability category termed dysrationalia--the inability to think and behave rationally despite adequate intelligence. The coining this term (which was called an "epistemological bender" by one commentator, see Metcalf, 1998) served as a critique of discrepancy definitions in learning disabilities and also served as a tool for exploring whether conceptual work could be done by differentiating intelligence and rationality in the manner described above.

I recognize that such a differentiation cuts against the grain of current trends in terminological practice. Specifically, it is the case that many prominent theorists depart from the distinction suggested above and prefer to conflate the terms rationality and intelligence in a manner more in line with folk psychology[2]. For example, Baron's (1985) use of the distinction between cognitive capacities and rational thinking dispositions is somewhat different from that exemplified in the concept of dysrationalia. He proposes that these dispositions be folded into our view of intelligence--that intelligence be made to encompass rationality. Perkins (1995; Perkins et al., 1993) likewise subsumes rationality within the construct of intelligence.

Similarly, Sternberg (1997a) explicitly defines intelligence in a manner that subsumes both epistemic and pragmatic rationality: "A more intelligent, adaptive person has achieved a higher degree of external correspondence and internal coherence in his or her knowledge base and belief structures. People think unintelligently to the extent that they make errors in achieving external correspondence or internal coherence. For example, in believing the gambler's fallacy, a person fails in achieving external correspondence; in touching a hot stove despite knowledge of the danger of doing so, a person fails in achieving internal coherence" (p. 1031). Thus, in appropriating external correspondence for the concept of intelligence Sternberg (1997a) encompasses epistemic rationality and the "hot stove" example is a clear case of pragmatic rationality. This intelligence conception quite thoroughly incorporates notions of rationality. Likewise, characterizations of intelligent behavior as that which helps us achieve our goals or that which helps us to adapt to the environment (Sternberg & Detterman, 1986) are conceptualizing intelligence as something overlapping with rationality, even if actual operationalizations of the concept do not reflect this.

All of these theorists have made progress with their conflated conceptions of intelligence and rationality, and thus my program of differentiating the two should not be seen as a replacing these efforts. There is more than one way to carve this particularly complex part of nature at its joints and each of the parsings has various strengths and weaknesses. I have alluded to some of these tradeoffs previously (Stanovich, 1993, 1994). For example, the conflated definition perhaps serves as a better platform for a critique of the properties of current IQ tests (a critique which I am in some sympathy with, see Stanovich, 1991, 1994). Allowing intelligence to subsume rationality highlights the fact that we <u>cannot</u> identify current IQ tests with the concept intelligence defined in this sense. In contrast, from the standpoint of a nonconflated definition, the tests--taken as omnibus indicators of overall functioning at the algorithmic level of analysis only--are less problematic first approximations.

Accord with vernacular usage (Neisser, 1979; Sternberg,, 1985; Sternberg, Conway, Ketron, & Bernstein, 1981) might be deemed another advantage of the conflated definition, although as I have discussed previously (Stanovich, 1990) convergence with folk psychology is very much a two-edged sword (Churchland, 1979; Churchland & Churchland, 1998). For example, the disability of dysrationalia disappears under this view which many may view as a virtue. But there is a cost to this disappearance. The cost is that we can no longer explain the "strangeness" of the notion of smart people acting dumb. Instead, we must accept the implication of the conflated view--smart people acting dumb really aren't as smart as we thought they were! Rational behavior is part of intelligence under the conflated view. What my view says is irrationality in the face of intelligence (dysrationalia) the conflated view says is impeached intelligence.

Despite some drawbacks, I have explored an alternative parsing of the psychological concepts in this chapter in order to give its advantages a fuller airing. First, as I have argued, the parsing I have argued for dissolves the somewhat paradoxical connotations of the "smart but acting dumb" phrase. Even more importantly, there are other important issues raised and at least partially answered by my distinctions. In the remainder of the chapter, I will highlight a few of these.

Individual Differences at the Intentional Level of Analysis

If we conflate intelligence and rationality in discussions of individual differences we lose the ability to address an issue of immense interest in philosophy and cognitive science: whether there can be actual (as opposed to apparent) variation in intentional-level psychologies. As discussed in Stanovich (1999) there are three powerful traditions in philosophy that argue against this possibility. Arguments from charity (Dennett, 1987; Quine, 1960; Stein, 1996; Stich, 1990), from reflective equilibrium (Cohen, 1981; Stein, 1996; Stich, 1990), and from evolution (Dennett, 1987; Stich, 1990) have famously claimed to have demonstrated uniformly optimal functioning of intentional-level psychologies in human beings.

I think all of these arguments are mistaken (Stanovich, 1999) but the more important point is that in order to produce empirical data relevant to the issue we need to clearly demarcate concepts at the intentional level.

Finding nonartifactual variation in a conflated notion of intelligence obscures the critical issue of whether the individual differences are best understood as arising from variation in algorithmic-level or intentional-level functioning. In contrast, by taking the demarcation as fundamental, I believe that the work summarized above (see Stanovich, 1999) has demonstrated that some smart people do a lot of dumb things and some don't--and that this is an indication of variation in intentional-level psychologies, variation in degrees of rationality that some philosophers have denied (see Cohen, 1981; Stanovich, 1999; Stein, 1996; Stich, 1990).

Fostering Actively Open-Minded Thinking: The Normative Issue

I also believe that drawing the intelligence/rationality distinction helps to provide a needed educational rationale for attempts to foster critical thinking. 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 (1993b) 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, 1989, 1993). This of course implies a normative model of what we mean by better thinking (Baron, Badgio, & Gaskins, 1986; Haslam & Baron, 1994).

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, 1992) 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, 1994). 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 the former and are essential for the latter. But at least in principle we could imagine a person with excellent epistemic rationality (their degree of confidence in propositions being well calibrated to the available evidence relevant to the proposition) and optimal practical rationality (they optimally satisfy their desires given their beliefs) who was not actively open-minded. We might still want to mold such an individual's dispositions in the direction of open-mindedness for the sake of society as a whole, but 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 if--whatever they were--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 thinking dispositions make the individual a more rational person (Baron, 1985, 1988, 1993b; Stanovich, 1994). But that puts a burden of proof upon the shoulders of advocates of such educational interventions. They must show that thinking dispositions are associated with the responses and thought patterns that are considered normative (and that the association is causal). This is precisely the empirical evidence that we (Stanovich, 1999; Stanovich & West, 1997, 1998, 2000) and other investigators (Kardash & Scholes, 1996; Klaczynski et al., 1997; Kuhn, 1991, 1993, 1996; Schaller, Boyd, Yohannes, & O'Brien, 1995; Schommer, 1990, 1994; Smith & Levin, 1996) have begun to compile. Although the trends are sometimes modest, there has been a consistent tendency for people who are high in actively open-minded thinking to give the normative response on hypothesis testing and reasoning tasks, to avoid belief bias in their reasoning, and to properly calibrate their beliefs to the state of the evidence. Thus, the field is beginning to develop a normatively justified foundation for an emphasis on thinking dispositions.

The Paradoxical Relation Between Rationality and Emotion

Identifying rationality with the intentional level of analysis in cognitive science and intelligence with the algorithmic level can potentially help to dissolve another seeming paradox--the disconnect between the folk theory of emotions and conceptions of the emotions in cognitive science. In folk psychology the emotions are ostensibly the cause of irrationality. Quintessentially, they are thought to interfere with rational thought. Yet despite the fact that folk psychology assigns them a disruptive role, most conceptions of emotions in cognitive science stress the adaptive regulatory powers of the emotions. For example, in their discussion of the rationality of emotions, Johnson-Laird and Oatley (1992; see Oatley, 1992) conceptualized emotions as interrupt signals supporting goal achievement. They see emotions as intentional-level constructs that are particularly important in the characterization of systems whose behavior is governed by neither fixed action patterns nor impeccable rationality. Other cognitive scientists concur in this view (see Damasio, 1994; de Sousa, 1987). The basic idea is that emotions serve to stop the combinatorial explosion of possibilities that would occur if an intelligent system tried to calculate the utility of all possible future outcomes. Emotions are thought to constrain the possibilities to a manageable number based on somatic markers (see Damasio, 1994) stored from similar situations in the past.

How are we to square this view with the folk psychological notion of the emotions as the enemies of reason? One potential resolution may reside in two-process models of cognitive activity that have been proposed by numerous investigators in the last two decades. These theories propose two structured cognitive systems with separable goal structures and separate algorithmic mechanisms to implement the goal structures. The details and terminology of these models differ, but they all share a family resemblance and the specific differences are not material for the present discussion. The dual-process terms of several major theorists are presented in Table 2. In order to emphasize the prototypical view that is adopted here, the two systems have simply been generically labeled System 1 and System 2. The key differences in the properties of the two systems are listed next. System 1 is characterized as automatic, heuristic-based, and relatively undemanding of computational capacity. Thus, it conjoins properties of automaticity and heuristic processing as these constructs have been variously discussed in the literature. System 2 conjoins the various characteristics that have been viewed as typifying controlled processing. At the algorithmic level, System 2 encompasses the processes of analytic intelligence that have traditionally been studied in psychometric work and that have been examined by information processing theorists trying to uncover the computational components underlying psychometric intelligence. At the intentional level, the goal structure of System 1 has been determined largely by evolutionary adaptation, whereas the goal structure of System 2 is more flexible and reflects ongoing goal evaluation at the personal level as an individual is shaped by environmental experience (see Stanovich, 1999).

Insert Table 2 about here

The work of Pollock (1991) is particularly relevant to the present discussion of the role of the emotions. In his view, heavily influenced by work in artificial intelligence, System 1 is composed of Q&I (quick & inflexible) modules that perform specific computations. System 2 processes are grouped under the term intellection in his model and refer to all explicit reasoning in the service of theoretical or practical rationality: "The advantage of Q&I modules is speed. The advantage of intellection, on the other hand, is extreme flexibility. It seems that it can in principle deal with any kind of situation, but it is slow" (p. 192).

As an example, Pollock (1991) mentions the Q&I trajectory module that predicts the movement path of objects in motion. The Q&I module for this computation is quite accurate, but it relies on certain assumptions about the structure of the world. When these assumptions are violated, then the module must be overridden by System 2 processing. So when a baseball approaches a telephone pole "we had best wait until it ricochets before predicting its trajectory. Our built-in trajectory module cannot handle this situation accurately, so we use intellection to temporarily override it until the situation becomes one that can be handled accurately by the trajectory module" (p. 191). Pollock (1991) stresses however that Q&I modules do not just operate in the domains of movement and perception but instead that "everyday inductive and probabilistic inference is carried out by Q&I modules" (p. 191). Indeed, Stanovich (1999; Stanovich & West, 2000) stressed the importance of the override function of System 2 in explaining individual differences in rational thought.

Importantly, Pollock (1995) conceptualizes emotions within his cognitive architecture in a manner that helps to dissolve the emotion/rationality paradox described above. In Pollock's (1995) model, emotions are conceived as Q&I modules for practical reasoning. As examples, Pollock (1995) notes that "being afraid of tigers initiates quick avoidance responses without our having to think about it--a very useful reaction for anyone who is likely to encounter tigers unexpectedly. Embarrassment, indignation, and so forth, may similarly be practical Q&I modules whose purpose is to supplement explicit practical reasoning in social situations. This provides a computational role for these emotions and throws light on why humans are subject to them" (p. 11).

Pollock's (1991, 1995) view is consistent with that of Johnson-Laird and Oatley (1992) and offers an explanation of the seeming discontinuity between the folk psychological view of the relation between emotions and rationality and the view of modern cognitive science. The key insight is that if we view emotions as Q&I modules for practical reasoning there are two ways in which the rational regulation of behavior could go wrong[3]. The two ways might be termed module failure and override failure, respectively. First, Q&I emotion modules might be missing or might malfunction. In this case, the automatic and rapid regulation of goals is absent and System 2 is faced with a combinatorial explosion of possibilities because the constraining function of the emotions is missing. A module failure of this type represents a case where there is not too much emotion but instead too little.

The second way that behavioral regulation can go awry has the opposite properties. It is a situation analogous to Pollock's (1995) trajectory example. Here, the Q&I module has fired but it happens to be one of those instances where the module's output is inappropriate and needs to be overridden by the controlled processing of System 2.

Behavioral regulation is suboptimal when the System 2 override function does not work properly. In this situation, the emotions of the Q&I practical reasoning module are too pervasive and unmodifiable. The problem in cases of override failure is indeed a problem of too much emotion, rather than too little.

It is clear that the folk psychological notion of the emotion/rationality relationship refers to the latter situation--failure to override System 1 Q&I modules for practical reasoning. This leads to the folk psychological cliche that emotion interferes with rational thought. But what folk psychology leaves out is irrationality of the first type--and here the emotions play the opposite role. It is their absence that is the problem. Behavioral regulation is not aided by crude but effective emotional signals that help to prioritize goals for subsequent action.

Folk psychology is thus incomplete in the sense that it recognizes override-based irrationality but not irrationality due to emotion module failure. Several architectures in cognitive science would, in contrast, recognize both possibilities (Damasio, 1994; Johnson-Laird & Oatley, 1992; Oatley, 1992; Pennington & Ozonoff, 1996; Pollock, 1991, 1995; Stanovich, 1999). More importantly, there is empirical evidence for rationality failures of the two different types. Dorsolateral prefrontal damage has been associated with executive functioning difficulties (and/or working memory difficulties) that can be interpreted as the failure of System 2 to override automatized processes being executed by System 1 (Duncan, Emslie, Williams, Johnson, & Freer, 1996; Kimberg, D'Esposito, & Farah, 1998; Kolb & Wilshaw, 1990; McCarthy & Warrington, 1990; Shallice, 1988). In contrast, ventromedial damage to the prefrontal cortex has been associated with problems in behavioral regulation that are accompanied by affective disruption (Bechara, Damasio, Damasio, & Anderson, 1994;

Bechara, Damasio, Tranel, & Damasio, 1997; Damasio, 1994). Difficulties of the former but not the latter kind are associated with lowered intelligence (Damasio, 1994; Duncan et al., 1996)--consistent with the association of System 2 with psychometric intelligence (see Table 2) and the relative independence of System 1 and algorithmic computational capacity of the type measured by IQ tests.

In summary, our developing understanding of the relation between emotion and rationality might provide another instance where cognitive science could well help to shape folk psychology in the direction of more accurate conceptions of human mental life (Bruner, 1990).

Epistemic Irrationality in the Face of Substantial Computational Power

It is not just practical rationality that can become dissociated from algorithmic efficiency. Epistemic rationality can also display marked dissociations. In fact, there is no dearth of examples of smart people believing ridiculous things--an indication that aspects of epistemic rationality (the proportional calibration of belief to evidence) can go awry. Studies of leading Holocaust deniers (see Lipstadt, 1994; Shermer, 1997) have revealed that their ranks contain the holder of a Master's degree from Indiana University in European history, the author of several well-known biographies of World War II figures, a professor of literature at the University of Lyon, an author of textbooks used in Ivy League universities, a professor of English at the University of Scranton, a professor at Northwestern University, and the list goes on (see Lipstadt, 1994).

A cognitive science that demarcates the intentional level also promises to throw some light on the puzzling phenomenon of epistemic irrationality coexisting with substantial cognitive power--the educated Holocaust deniers studied by Lipstadt (1994), creationists who are physical scientists, and many similar examples (Shermer, 1997; Stanovich, 1993). Philosopher

Hilary Kornblith (1993) provides one form of the argument that can unlock this seeming puzzle. In discussing the phenomenon of belief perseverance: "mistaken beliefs will, as a result of belief perseverance, taint our perception of new data. By the same token, however, belief perseverance will serve to color our perception of new data when our preexisting beliefs are accurate....If, overall, our belief-generating mechanisms give us a fairly accurate picture of the world, then the phenomenon of belief perseverance may do more to inform our understanding than it does to distort it" (p. 105).

This argument--that in a natural ecology where most of our prior beliefs are true, projecting our beliefs on to new data will lead to faster accumulation of knowledge--I have termed the knowledge projection argument (Stanovich, 1999), and it reappears in a remarkably diverse set of contexts throughout the reasoning and decision making literature. For example, Koehler (1993) demonstrated that scientists' prior beliefs about a hypothesis influenced their judgments of evidence quality. In a Bayesian analysis of whether this evaluation tendency could ever be normatively justified, Koehler (1993) found that under certain conditions it could. One of those conditions was that the prior hypotheses influencing evidence evaluation were more likely than not to be true. When evidence is evaluated with reference to a pool of hypotheses that are largely true, that evidence will lead to belief convergence faster if the prior beliefs do influence evidence evaluation--another version of the knowledge projection argument.

Evans, Over, and Manktelow (1993) rely on a variant of the knowledge projection argument when considering the normative status of belief bias in syllogistic reasoning. They consider the status of selective scrutiny explanations of the belief bias phenomenon. Such theories posit that subjects accept conclusions that are believable without engaging in logical reasoning at all. Only when faced with unbelievable conclusions do subjects engage in logical reasoning about the premises. Evans et al. (1993) consider whether such a processing strategy could be rational in the sense of serving to achieve the person's goals, and they conclude that it could. They argue that any adult is likely to hold a large number of true beliefs that are interconnected in complex ways. Because single-belief revision has interactive effects on the rest of the belief network, it may be computationally costly. Evans et al. (1993) argue that under such conditions it is quite right that conclusions that contradict one's beliefs "should be subjected to the closest possible scrutiny and refuted if at all possible" (p. 174). Again, the argument works when the selective scrutiny mechanism is applied using a subset of beliefs that are largely true in the domain to which the scrutiny strategy is being applied.

Finally, Alloy and Tabachnik (1984) echo the knowledge projection argument in their review of the covariation detection literature on humans and other animals: "when individuals' expectations accurately reflect the contingencies encountered in their natural environments... it is not irrational for them to assimilate incoming information about covariation between events to these expectations....Because covariation information provided in an experiment may represent only one piece of conflicting evidence against the background of the large body of data about event covariations summarized by an expectation, it would be normatively appropriate for organisms to weight their expectations more heavily than situational information in the covariation judgment process" (p. 140). Of course, Alloy and Tabachnik (1984) emphasize that we must project from a largely <u>accurate</u> set of beliefs in order to obtain the benefit of knowledge projection. In a sea of inaccurate beliefs, the situation is quite different.

And herein lies the key to understanding the creationist or Holocaust denier.

The caveat here is critical: When the subset of beliefs that the individual is projecting contains substantial false information, selective scrutiny will delay the assimilation of the correct information. This caveat creates the possibility of observing a so-called "Matthew Effect"--a cumulative advantage phenomenon--in the acquisition of knowledge (Stanovich, 1986). Walberg (Walberg et al., 1984; Walberg & Tsai, 1983), following Merton (1968), dubbed cumulative advantage effects in education "Matthew effects," after the Gospel according to Matthew: "For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath" (XXV:29). In the educational literature, the term springs from findings that individuals who have advantageous early educational experiences are able to utilize new educational experiences more efficiently and thus increase their advantage. How might the knowledge projection process lead to Matthew Effects in knowledge acquisition? Image two scientists, A and B, working in domain X. The bulk of hypotheses in domain X held by scientist A are true and the bulk of hypotheses in domain X held by scientist B are false. Imagine that they both then begin to project those prior beliefs on the same new evidence in the manner demonstrated experimentally by Koehler (1993)--with stronger tendencies to undermine the evidence when it contradicted prior belief. It is clear that scientist A--who already exceeds B in number of true beliefs--will increase that advantage as new data comes in.

The knowledge projection tendency, efficacious in the aggregate, may have the effect of isolating certain individuals on "islands of false beliefs" from which--because of the knowledge projection tendency--they are unable to escape. In short, there may be a type of knowledge isolation effect when projection is used in particularly ill-suited circumstances. Thus, knowledge projection, which in the aggregate might lead to more rapid induction of new true beliefs, may be a trap in cases where people, in effect, keep reaching into a bag of beliefs which are largely false, using these beliefs to structure their evaluation of evidence, and hence more quickly adding incorrect beliefs to the bag for further projection.

Knowledge projection from an island of false beliefs might explain the phenomenon of otherwise intelligence people who get caught in a domain-specific web of falsity and because of projection tendencies cannot escape (e.g., the otherwise competent physical scientists who believe in creationism). Indeed, such individuals often use their considerable computational power to rationalize their beliefs and to ward off the arguments of skeptics (Evans, 1996; Evans & Wason, 1976; Margolis, 1987; Nisbett & Wilson, 1977; Wason, 1969). The cognitive machinery recruited to aid in knowledge projection might be extremely potent in individuals high in cognitive capacity--but when the projection occurs from an island of false belief it merely results in a belief network even more divergent from that of individuals not engaged in such projection or with less computational power.

Further research is needed to examine whether such Matthew Effects and knowledge isolation effects can be documented. Nevertheless, it provides a statistical rationale for the presence of such a bias, because across individuals--and across beliefs within an individual--most of what is believed is true. Thus, on an overall statistical basis, knowledge projection may well increase the rate of acquisition of true beliefs. But this does not prevent particular individuals with particularly ill-formed initial beliefs from projecting them and developing beliefs which are even less in correspondence with reality. Neither does it prevent an individual (with an otherwise generally accurate belief network) from getting caught on an island of false beliefs with respect to a particular domain, projecting those beliefs, and with time developing even more bizarre theories about this domain. These effects might explain how some individuals could have their beliefs detached from reality in ways so extreme that an attribution of irrationality would seem justified. Such a case would be an example of a generally efficacious mechanism resulting in seriously suboptimal belief structures. Knowledge projection is thus a mechanism that could be generally normative in a statistical sense but still be the cause of a minority of actions and beliefs that are seriously irrational.

Notes

Preparation of this chapter was supported by a grant from the Social Sciences and Humanities Research Council of Canada to Keith E. Stanovich.

[1] Bruner (1986, 1990) has argued that researchers should display a greater awareness of their influence on folk psychology because the intuitive psychologies of the layperson provide the motivation for social policies. Rationality assumptions of various types form an important part of folk concepts of the nature of human cognition and, as Bruner (1990) notes, "it is through folk psychology that people anticipate and judge one another, draw conclusions about the worthwhileness of their lives, and so on. Its power over human mental functioning and human life is that it provides the very means by which culture shapes human beings to its requirements" (p. 15).

[2] The terms must be somewhat conflated or else the smart but acting dumb phrasing would not sound strange at all.

[3] Obviously there may be more than two. I am focusing on one particular contrast here.

References

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

Ackerman, P. L., Kyllonen, P. , & Roberts, R. (Eds.) (1999). The future of learning and individual differences research: Processes, traits, and content. Washington, DC: American Psychological Association.

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.

Alloy, L. B., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. Psychological Review, 91, 112-149.

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

Anderson, J. R. (1991). Is human cognition adaptive? Behavioral and Brain Sciences, 14, 471-517.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. Psychological Bulletin, 110, 486-498.

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

Arkes, H., & Hammond, K. (Eds.) (1986). Judgment and decision making. Cambridge, England: Cambridge University Press.

Audi, R. (1993). Action, intention, and reason. Ithaca, NY: Cornell University Press.

Babad, E., & Katz, Y. (1991). Wishful thinking--Against all odds. Journal of Applied Social Psychology, 21, 1921-1938.

Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. Developmental Psychology, 23, 611-626.

Baron, J. (1985). Rationality and intelligence. Cambridge: Cambridge University Press.

Baron, J. (1988). Thinking and deciding. Cambridge, England: Cambridge University Press.

Baron, J. (1991). Beliefs about thinking. In J. Voss,D. Perkins, & J. Segal (Eds.), Informal reasoning and education (pp. 169-186). Hillsdale, NJ: Erlbaum.

Baron, J. (1993a). Morality and rational choice. Dordrecht: Kluwer.

Baron, J. (1993b). 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. (1998). Judgment misguided: Intuition and error in public decision making. New York: Oxford University Press.

Baron, J., Badgio, P. C., & Gaskins, I. W. (1986). Cognitive style and its improvement: A normative approach. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence (Vol. 3, pp. 173-220). Hillsdale, NJ: Erlbaum.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. Journal of Personality and Social Psychology, 54, 569-579.

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

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997, Feb. 26). Deciding advantageously before knowing the advantageous strategy. Science, 275, 1293-1295.

Belsky, G. (1995, July). Why smart people make major money mistakes. Money Magazine, pp. 76-85.

Belsky, G., & Gilovich, T. (1999). Why smart people make big money mistakes--And how to correct them: Lessons from the new science of behavioral economics. New York: Simon & Schuster.

Bratman, M. E., Israel, D. J., & Pollack, M. E. (1991). Plans and resource-bounded practical reasoning. In J. Cummins & J. Pollock (Eds.), Philosophy and AI: Essays at the interface (pp. 7-22). Cambridge, MA: MIT Press.

Brenner, L. A., Koehler, D. J., & Tversky, A. (1996). On the evaluation of one-sided evidence. Journal of Behavioral Decision Making, 9, 59-70.

Bruner, J. (1986). Actual minds, possible worlds. Cambridge, MA: Harvard University Press.

Bruner, J. (1990). Acts of meaning. Cambridge, MA: Harvard University Press.

Buss, D. M. (1991). Evolutionary personality psychology. Annual Review of Psychology, 42, 459-491.

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.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. Psychological Review, 97, 404-431.

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

Carroll, J. B. (1997). Psychometrics, intelligence, and public perception. Intelligence, 24, 25-52.

Cederblom, J. (1989). Willingness to reason and the identification of the self. In E. Maimon, D. Nodine, & O'Conner (Eds.), Thinking, reasoning, and writing (pp. 147-159). New York: Longman.

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

Christensen, S. M., & Turner, D. R. (Eds.). (1993). Folk psychology and the philosophy of mind. Hillsdale, NJ: Lawrence Erlbaum Associates.

Churchland, P. M. (1979). Scientific realism and the plasticity of mind. Cambridge: Cambridge University Press.

Churchland, P. M., & Churchland, P. S. (1998). On the contrary: Critical essays, 1987-1997. Cambridge, MA: MIT Press.

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

Costa, P. T., & McCrae, R. R. (1992). Revised NEO personality inventory. Odessa, FL: Psychological Assessment Resources.

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

Davies, M., & Stone, T. (Ed.). (1995). Folk psychology. Oxford: Blackwell.

Dawes, R. M. (1988). Rational choice in an uncertain world. San Diego, CA: Harcourt, Brace Jovanovich.

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

Dawkins, R. (1998). Unweaving the rainbow. Boston: Houghton Mifflin.

Deary, I. J., & Stough, C. (1996). Intelligence and inspection time.

American Psychologist, 51, 599-608.

Dennett, D. C. (1978). Brainstorms: Philosophical essays on mind and psychology. Cambridge, MA: MIT Press.

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

de Sousa, R. (1987). The rationality of emotion. Cambridge, MA: MIT Press.

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

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. Journal of Experimental Psychology: General, 128, 309-331.

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

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. American Psychologist, 49, 709-724.

Epstein, S., & Meier, P. (1989). Constructive thinking: A broad coping variable with specific components. Journal of Personality and Social Psychology, 57, 332-350.

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

Evans, J. St. B. T. (1989). Bias in human reasoning: Causes and consequences. London: Erlbaum Associates.

Evans, J. St. B. T. (1996). Deciding before you think: Relevance and reasoning in the selection task. British Journal of Psychology, 87, 223-240.

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., 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, England: Psychology Press.

Evans, J. St. B. T., Over, D. E., & Manktelow, K. (1993). Reasoning, decision making and rationality. Cognition, 49, 165-187.

Evans, J. St. B. T., & Wason, P. C. (1976). Rationalization in a reasoning task. British Journal of Psychology, 67, 479-486.

Facione, P. (1992). California Critical Thinking Dispositions Inventory. La Cruz, CA: California Academic Press.

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

Foley, R. (1991). Rationality, belief, and commitment. Synthese, 89, 365-392.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. Cognitive Psychology, 18, 253-292.

Fridson, M. S. (1993). Investment illusions. New York: Wiley.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence. Psychological Science, 7, 237-241.

Gibbard, A. (1990). Wise choices, apt feelings: A theory of normative judgment. Cambridge, MA: Harvard University Press.

Gilovich, T.(1991). How we know what isn't so. New York: Free Press.

Goff, M., & Ackerman, P. L. (1992). Personality-intelligence relations: Assessment of typical intellectual engagement. Journal of Educational

Psychology, 84, 537-552.

Goldberg, L. R. (1992). The social psychology of personality. Psychological Inquiry, 3, 89-94.

Greenwood, J. D. (Ed.). (1991). The future of folk psychology: Intentionality and cognitive science. Cambridge: Cambridge University Press.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. Cognitive Psychology, 24, 411-435.

Harman, G. (1995). Rationality. In E. E. Smith & D. N. Osherson (Eds.), Thinking (Vol. 3) (pp. 175-211). Cambridge, MA: The MIT Press.

Haslam, N., & Baron, J. (1994). Intelligence, personality, and prudence. In R. J. Sternberg & P. Ruzgis (Eds.), Personality and intelligence (pp. 32-58). Cambridge: Cambridge University Press.

Horgan, T., & Tienson, J. (1993). Levels of description in nonclassical cognitive science. In C. Hookway & D. Peterson (Eds.), Philosophy and cognitive science (pp. 159-188). Cambridge: Cambridge University Press.

Hunt, E. (1978). Mechanics of verbal ability. Psychological Review, 85, 109-30.

Hunt, E. (1987). The next word on verbal ability. In P. A. Vernon (Ed.), Speed of information-processing and intelligence (pp. 347-392). Norwood, NJ: Ablex.

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

Jepson, C., Krantz, D., & Nisbett, R. (1983). Inductive reasoning: Competence or skill? Behavioral and Brain Sciences, 6, 494-501.

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

Johnson-Laird, P., & Oatley, K. (1992). Basic emotions, rationality, and folk theory. Cognition and Emotion, 6, 201-223.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.

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

Kardash, C. M., & Scholes, R. J. (1996). Effects of pre-existing beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues. Journal of Educational Psychology, 88, 260-271.

Keating, D. P. (1990). Charting pathways to the development of expertise. Educational Psychologist, 25, 243-267.

Kimberg, D. Y., D'Esposito, M., & Farah, M. J. (1998). Cognitive functions in the prefrontal cortex--working memory and executive control. Current Directions in Psychological Science, 6, 185-192.

King, P. M., & Kitchener, K. S. (1994). Developing reflective judgment. San Francisco: Jossey-Bass.

Kitchener, K., & Fischer, K. W. (1990). A skill approach to the development of reflective thinking. In D. Kuhn (Eds.), Developmental perspectives on teaching and learning thinking skills (pp. 48-62). Basil: Karger.

Kitcher, P. (1993). The advancement of science. New York: Oxford University Press.

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. (1996). Self-serving influences on adolescents' evaluations of belief-relevant evidence. Journal of

Experimental Child Psychology, 62, 317-339.

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., & Narasimham, G. (1998). Development of scientific reasoning biases: Cognitive versus ego-protective explanations. Developmental Psychology, 34, 175-187.

Klein, G. (1998). Sources of power: How people make decisions. Cambridge, MA: MIT Press.

Kleindorfer, P. R., Kunreuther, H. C., & Schoemaker, P. J. H. (1993). Decision sciences: An integrative perspective. Cambridge: Cambridge University Press.

Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. Organizational Behavior and Human Decision Processes, 56, 28-55.

Kolb, B., & Whishaw, I. (1990). Fundamentals of human neuropsychology (Third Ed.). New York: W. H. Freeman.

Kornblith, H. (1993). Inductive inference and its natural ground. Cambridge, MA: MIT University Press.

Kruglanski, A. W. (1990). Lay epistemics theory in social-cognitive psychology. Psychological Inquiry, 1, 181-197.

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

Kuhn, D. (1991). The skills of argument. Cambridge: Cambridge University Press.

Kuhn, D. (1992). Thinking as argument. Harvard Educational Review, 62, 155-178.

Kuhn, D. (1993). Connecting scientific and informal reasoning. Merrill-

Palmer Quarterly, 38, 74-103.

Kuhn, D. (1996). Is good thinking scientific thinking? In D. R. Olson & N. Torrance (Eds.), Modes of thought: Explorations in culture and cognition (pp. 261-281). New York: Cambridge University Press.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). The development of scientific thinking skills. San Diego: Academic Press.

Kunda, Z. (1990). The case for motivated reasoning. Psychological Bulletin, 108, 480-498.

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

Levelt, W. (1995). Chapters of psychology. In R. L. Solso & D. W. Massaro (Eds.), The science of the mind: 2001 and beyond (pp. 184-202). New York: Oxford University Press.

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

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. Journal of Experimental Psychology: Human Learning and Memory, 4, 551–578.

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

Lipstadt, D. (1994). Denying the Holocaust. New York: Plume.

Lohman, D. F. (1989). Human intelligence: An introduction to advances in theory and research. Review of Educational Research, 59, 333-373.

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

Margolis, H. (1996). Dealing with risk. Chicago: University of Chicago Press.

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

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

McNeil, B., Pauker, S., Sox, H., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. New England Journal of Medicine, 306, 1259-1262.

Merton, R. K. (1968). The Matthew effect in science. Science, 159, 59-63.

Messick, S. (1984). The nature of cognitive styles: Problems and promise in educational practice. Educational Psychologist, 19, 59-74.

Messick, S. (1994). The matter of style: Manifestations of personality in cognition, learning, and teaching. Educational Psychologist, 29, 121-136.

Metcalf, S. D. (1998, March). Attention deficits. Lingua Franca, 8(2), 60-64.

Moshman, D. (1994). Reasoning, metareasoning, and the promotion of rationality. In A. Demetriou & A. Efklides (Eds.), Intelligence, mind, and reasoning: Structure and development (pp. 135-150). Amsterdam: Elsevier.

Nathanson, S. (1994). The ideal of rationality. Chicago: Open Court.

Navon, D. (1989a). The importance of being visible: On the role of attention in a mind viewed as an anarchic intelligence system: I. Basic tenets. European Journal of Cognitive Psychology, 1, 191-213.

Navon, D. (1989b). The importance of being visible: On the role of attention in a mind viewed as an anarchic intelligence system: II. Application to the field of attention. European Journal of Cognitive Psychology, 1, 215-238.

Neisser, U. (1979). The concept of intelligence. In R. J. Sternberg & D. K. Detterman (Eds.), Human intelligence (pp. 179-189). Norwood, NJ: Ablex.

Newell, A. (1982). The knowledge level. Artificial Intelligence, 18, 87-127.

Newell, A. (1990). Unified theories of cognition. Cambridge, MA: Harvard University Press.

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

Nickerson, R. S. (1988). On improving thinking through instruction. In E. Z. Rothkopf (Eds.), Review of Research in Education (Vol. 15) (pp. 3-57). Washington, DC: American Educational Research Association.

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

Nisbett, L., & Ross, L. (1980). Human inference: Strategies and shortcomings of social judgment. Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A. (1976). Memory and attention: An introduction to human information processing, (2nd Ed.), New York: John Wiley & Sons, Inc.

Norris, S. P. (1992). Testing for the disposition to think critically. Informal Logic, 14, 157-164.

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

Nozick, R. (1993). The nature of rationality. Princeton, NJ: Princeton University Press.

Oakhill, J., Johnson-Laird, P. N., & Garnham, A. (1989). Believability and syllogistic reasoning. Cognition, 31, 117-140.

Oaksford, M., & Chater, N. (1993). Reasoning theories and bounded rationality. In K. Manktelow & D. Over (Eds.), Rationality: Psychological and philosophical perspectives (pp. 31-60). London: Routledge.

Oaksford, M., & Chater, N. (1995). Theories of reasoning and the computational explanation of everyday inference. Thinking and

Reasoning, 1, 121-152.

Oatley, K. (1992). Best laid schemes: The psychology of emotions. Cambridge: Cambridge University Press.

Osherson, D. N. (1995). Probability judgment. In E. E. Smith & D. N. Osherson (Eds.), Thinking (Vol. 3) (pp. 35-75). Cambridge, MA: The MIT Press.

Paul, R. W. (1984). Critical thinking: Fundamental to education for a free society. Educational Leadership, 42(1), 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.

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

Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. Journal of Educational Psychology, 77, 562-571.

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

Perkins, D. N., Farady, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In J. Voss,D. Perkins, & J. Segal (Eds.), Informal reasoning and education (pp. 83-105). Hillsdale, NJ: Erlbaum.

Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. Merrill-Palmer Quarterly, 39, 1-21.

Piattelli-Palmarini, M. (1994). Inevitable illusions: How mistakes of reason rule our minds. New York: John Wiley.

Pintrich, P. R., Marx, R., & Boyle, R. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. Review of Educational

Research, 63, 167-199.

Plous, S. (1993). The psychology of judgment and decision making. New York: McGraw-Hill.

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

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

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Eds.), Information processing and cognition: The Loyola Symposium (pp. 55-85). New York: Wiley.

Pylyshyn, Z. (1984). Computation and cognition. Cambridge, MA: MIT Press.

Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (pp. 297-340). San Diego: Academic Press.

Quine, W. (1960). Word and object. Cambridge, MA: MIT Press.

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

Redelmeier, D. A., & Tversky, A. (1990). Discrepancy between medical decisions for individual patients and for groups. New England Journal of Medicine, 322, 1162-1164.

Redelmeier, D. A., & Tversky, A. (1992). On the framing of multiple prospects. Psychological Science, 3, 191-193.

Rolfhus, E. L., & Ackerman, P. L. (1999). Assessing individual differences in knowledge: Knowledge, intelligence, and related skills. Journal of

Educational Psychology, 91, 511-526.

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.

Saks, M., & Kidd, R. (1980–1981). Human information processing and adjudication: Trial by heuristics. Law and Society Review, 15, 123–160.

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

Schaller, M., Boyd, C., Yohannes, J., & O'Brien, M. (1995). The prejudiced personality revisited: Personal need for structure and formation of erroneous stereotypes. Journal of Personality and Social Psychology, 68, 544-555.

Scheffler, I. (1991). In praise of the cognitive emotions. New York: Routledge.

Schoenfeld, A. H. (1983). Beyond the purely cognitive: Belief systems, social cognitions, and metacognitions as driving forces in intellectual performance. Cognitive Science, 7, 329-363.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. Journal of Educational Psychology, 82, 498-504.

Schommer, M. (1993). Epistemological development and academic performance among secondary students. Journal of Educational Psychology, 85, 406-411.

Schommer, M. (1994). Synthesizing epistemological belief research: Tentative understandings and provocative confusions. Educational Psychology Review, 6, 293-319.

Schrag, F. (1988). Thinking in school and society. New York: Routledge.

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

Shafir, E., & Tversky, A. (1995). Decision making. In E. E. Smith & D. N.

Osherson (Eds.), Thinking (Vol. 3) (pp. 77-100). Cambridge, MA: The MIT Press.

Shallice, T. (1988). From neuropsychology to mental structure. Cambridge: Cambridge University Press.

Shermer, M. (1997). Why people believe weird things. New York: WH Freeman.

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

Siegel, H. (1993). Not by skill alone: The centrality of character to critical thinking. Informal Logic, 15, 163-177.

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

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

Slusher, M. P., & Anderson, C. A. (1996). Using causal persuasive arguments to change beliefs and teach new information: The mediating role of explanation availability and evaluation bias in the acceptance of knowledge. Journal of Educational Psychology, 88, 110-122.

Smith, S. M., & Levin, I. P. (1996). Need for cognition and choice framing effects. Journal of Behavioral Decision Making, 9, 283-290.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. Reading Research Quarterly, 21, 360-407.

Stanovich, K. E. (1989). Implicit philosophies of mind: The dualism scale and its relationships with religiosity and belief in extrasensory perception. Journal of Psychology, 123, 5-23.

Stanovich, K. E. (1990). Concepts in developmental theories of reading skill: Cognitive resources, automaticity, and modularity. Developmental Review, 10, 72-100.

Stanovich, K. E. (1991). Discrepancy definitions of reading disability: Has intelligence led us astray? Reading Research Quarterly, 26, 7-29.

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

Stanovich, K. E. (1994). Reconceptualizing intelligence: Dysrationalia as an intuition pump. Educational Researcher, 23(4), 11-22.

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

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. (1998). Individual differences in rational thought. Journal of Experimental Psychology: General, 127, 161-188.

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

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

Sterelny, K. (1990). The representational theory of mind: An introduction. Oxford: Basil Blackwell.

Sternberg, R. J. (Ed.). (1982). Handbook of human intelligence. Cambridge: Cambridge University Press.

Sternberg, R. J. (1985). Implicit theories of intelligence, creativity, and wisdom. Journal of Personality and Social Psychology, 49, 607-627.

Sternberg, R. J. (1988). Mental self-government: A theory of intellectual styles and their development. Human Development, 31, 197-224.

Sternberg, R. J. (1989). Domain-generality versus domain-specificity: The life and impending death of a false dichotomy. Merrill-Palmer Quarterly, 35, 115-130.

Sternberg, R. J. (1997a). The concept of intelligence and its role in lifelong learning and success. American Psychologist, 52, 1030-1037.

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

Sternberg, R. J., Conway, B., Ketron, J., & Bernstein, M. (1981). People's conceptions of intelligence. Journal of Personality and Social Psychology, 41, 37-55.

Sternberg, R. J., & Detterman, D. K. (1986). What is intelligence? Norwood, NJ: Ablex.

Sternberg, R. J., & Ruzgis, P. (Eds.). (1994). Personality and intelligence. Cambridge: Cambridge University Press.

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

Stich, S. P. (1996). Deconstruction of the mind. New York: Oxford University Press.

Sutherland, S. (1992). Irrationality: The enemy within. London: Constable.

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

Thaler, R. H. (1992). The winner's curse: Paradoxes and anomalies of economic life. New York: Free Press.

Tobacyk, J., & Milford, G. (1983). Belief in paranormal phenomena. Journal of Personality and Social Psychology, 44, 1029-1037.

Tversky, A. (1996). Contrasting rational and psychological principles of choice. In R. Zeckhauser, R. Keeney, & J. Sebenius (Eds.), Wise choices (pp. 5-21). Boston, MA: Harvard Business School Press.

Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. Psychological Science, 3, 305-309.

Vernon, P. A. (1991). The use of biological measures to estimate behavioral intelligence. Educational Psychologist, 25, 293-304.

Vernon, P. A. (1993). Biological approaches to the study of human intelligence. Norwood, NJ: Ablex.

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

Walberg, H. J., & Tsai, S. (1983). Matthew effects in education. American Educational Research Journal, 20, 359-373.

Walberg, H. J., Strykowski, B. F., Rovai, E., & Hung, S. S. (1984). Exceptional performance. Review of Educational Research, 54, 87-112.

Wason, P. C. (1969). Regression in reasoning? British Journal of Psychology, 60, 471-480.

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

Willis, C. (1990, June). The ten mistakes to avoid with your money. Money Magazine, pp. 84–94.

Yates, J. F. (Ed.). (1992). Risk-taking behavior. Chichester, England: John Wiley.

Table 1

<u>Different Levels of Cognitive Theory as Characterized by Several Investigators and in This Chapter</u>

Anderson	Marr	Newell	Dennett	This Chapter
Rational Level	Computational Level	Knowledge Level	Intentional Stance	Intentional Level
Algorithmic Level	Representation and Algorithm	Program Symbol Level	Design Stance	Algorithmic Level
Implementation Level		Register Transfer Level		
Biological Level	Hardware Implementation	Device	Physical Stance	Biological Level

Table 2

The Terms for the Two Systems Used by a Variety of Theorists and the Properties of Dual-Process Theories of Reason

	System 1	System 2
Dual-process theories:		
Sloman (1996)	Associative system	Rule-based system
Evans (1984, 1989)	Heuristic processing	Analytic processing
Evans & Over (1996)	Tacit thought processes	Explicit thought processes
Reber (1993)	Implicit cognition	Explicit learning
Levinson (1995)	Interactional intelligence	Analytic intelligence
Epstein (1994)	Experiential system	Rational system
Pollock (1991)	Quick & inflexible modules	Intellection
Klein (1998)	Recognition-primed decisions	Rational choice strategy
Johnson-Laird (1983)	Implicit inferences	Explicit inferences
Chaiken, Liberman, & Eagly (1989)	Heuristic processing	Systematic processing
Gibbard (1990)	Animal control system	Normative control system
Shiffrin & Schneider (1977)	Automatic processing	Controlled processing
Posner & Snyder (1975)	Automatic activation	Conscious processing system
Evans & Wason (1976)	Type 1 processes	Type 2 processes
Properties:	Associative	Rule-based
	Holistic	Analytic
	Automatic	Controlled
	Relatively undemanding of cognitive capacity	Demanding of cognitive capacity
	Relatively fast	Relatively slow
	Acquisition by biology, exposure, and personal experience	Acquisition by cultural and formal tuition
	Highly contextualized	Decontextualized
Goal structure	Largely genetically determined	Utility maximizing for the organism and constantly updated because of changes in environment
Type of intelligence indexed:	Interactional (conversational implicature)	Analytic (psychometric IQ)