On Motivated Reasoning and Self-Belief

Erik G. Helzer and David Dunning

Cornell University

Running Head: Motivated Reasoning and Self-Belief

In 2010, a team of political scientists published a study with a curious result. They invited roughly 200 residents of Eastern Iowa to take part in a hypothetical presidential primary election for either the Democratic or Republican Party. Participants read about various candidates, gathered information about each one, and then selected their tentative favorite. The research team then introduced an important wrinkle. For some participants, their favorite candidate continued to express ideologically consistent opinions, presumably opinions the participant liked. For others, the candidates expressed opinions that opposed the participants' stance about 25% of the time. For another group, the candidate almost completely reversed—taking positions that disagreed with the participant about 75% of the time (Redlawsk, Civettini, & Emmerson, 2010).

At the end of the study, participants were asked to offer a final decision about which candidate they preferred. Not surprisingly, participants whose candidates stayed uniformly true to participants' own views showed little change from their initial preferences. Also unsurprising, participants confronting candidates who ended up mostly disagreeing with them tended to significantly lower their initially positive impression of the candidate. But what about participants whose preferred candidates showed 25% disagreement? One might expect that participants would like these candidates less than the candidates who showed uniform agreement. But that is not what the research team found. Instead, participants confronting candidates displaying this level of disagreement rated them *more*, not less, favorably than they had when they offered their initial rating (Redlawsk et al., 2010). How could this be? How could disagreement lead to greater liking rather than its opposite?

What this study shows is that a little disagreement never got in the way of potentially vigorous *motivated reasoning*. Motivated reasoning refers to thought or analysis that is aimed at supporting some favored conclusion. People like to think, for example, that their future is a bright and benign one, and so may bend, massage, mold, select, or favor arguments and evidence that favor that conclusion over its opposite (Beckman, 1973; Willis, 1981; Wyer & Frey, 1983; Pyszczynski, Greeberg, & Holt, 1985; Kunda, 1990; Ditto & Lopez, 1992). In the Iowa study, one can see how people might engage in motivated reasoning after they had come to prefer a specific candidate. Perturbed by seemingly disagreeable "flaws" in their candidates, people may have reacted by explaining away those flaws or bolstering other aspects of the candidate's politics (Festinger, 1957). Indeed, the research team found presumed evidence of this, in that participants spent more time reading and thinking over disagreements with their favored candidates than they did agreements (Redlawsk et al., 2010).

Some Introductory Notes About Motivated Reasoning

If people engage in motivated reasoning to defend a hypothetical candidate in some fictional presidential primary in a researcher's lab, imagine how strongly they react when thinking about actual people for whom they have real feelings. Take, for example, how much people might engage in motivated reasoning when it is the *self* who is in question. That issue is the focus of the present chapter: How does motivated reasoning influence thinking about the self? And to what extent does it enhance or detract from people's understanding of themselves?

However, before beginning our discussion of motivated reasoning and self-knowledge, we must make a few introductory points. First, in a sense, all reasoning is motivated in some way. A mathematician is motivated to prove some theorem, but that is not the sense of motivation we focus on herein. The motivation spurring on the mathematician is simply the need to know, or curiosity, which has been shown to be an important driver of human thought

(Dunning, 2001; Kruglanski, Orehek, Dechesne, & Pierro, 2009). Similarly, people might be motivated by a desire to be accurate in their thinking, but that again is not what we specifically mean by motivated reasoning (Dunning, 2001). Further, people might be motivated to reach an impression of the world that is coherent and that contains no puzzling or discomforting contradictions. For several years, from the 1940s to the 1970s, social psychology emphasized this press toward cognitive consistency (for a review, see Abelson, Aronson, McGuire, Newcomb, Rosenberg, & Tannenbaum, 1968), and in fact it lay at the heart of the original formulation of both cognitive dissonance (Festinger, 1957) and balance (Heider, 1946) theories. But, this kind of motivation is, again, not the motivated reasoning discussed in this chapter.

What we mean, and what researchers typically mean, when we talk about motivated reasoning in contemporary psychological research is thought that is *directional*, i.e., thought that favors some pre-determined conclusion that the individual desires to reach (Kunda, 1990). This type of reasoning has many names: *wishful thinking*, *rationalization*, *self-affirmation* and *defensive processing*, just to name a few (Festinger, 1957; Taylor, 1983; Steele, 1988; Taylor & Brown, 1988; Ditto & Lopez, 1992). It can take on two forms. First, there may be some explicit conclusion in conscious deliberation that the individual favors reaching—like people convincing themselves (as much as the job interviewer) that they are the best person for the job (Kunda, 1987, 1990; Ditto & Lopez, 1992).

Second, there may be some background belief that the individual does not wish to contradict even if the belief is never made explicit. For example, much work has shown that people judge others in ways that reaffirm that they themselves are lovable and capable human beings. Even when the self is not explicitly mentioned, people are more likely to judge a target who shares their own strengths and weaknesses as capable, while a dissimilar other tends to be seen as less skilled (for reviews, see Dunning, 1999, 2007).

As such, motivated reasoning holds two general implications for self-knowledge. First, motivated reasoning leads to conclusions that contain a good deal of bias. The conclusions that people reach often lay some distance from objective truth or an impartial reading of the evidence. Second, motivated reasoning leads people to persist in believing conclusions about themselves far more than they should. Even when the evidence suggests that people should revise what they think about themselves, motivated reasoning allows them to cling to favored beliefs and attitudes.

We begin our discussion by describing some primary strategies people use to reason their way toward a desired conclusion. We will then turn to current controversies in the literature, focusing, for example, on whether motivated reasoning *actually* leads to inaccuracies in self-knowledge. We conclude with lingering questions that have yet to be satisfactorily answered about this topic, and thus require further theoretical and empirical work.

Strategies of Motivated Reasoning

The psychological literature is full of research documenting a wide array of strategies people use to travel closer to conclusions they wish to reach (for reviews, see Kunda, 1990; Baumeister & Newman, 1994; Dunning, 2001). We focus more narrowly on the strategies that have the greatest implications for self-knowledge and -understanding. The conclusions people wish to reach about themselves may influence how they process information about the world in any number of ways—some blatant, some subtle.

Motives Influence the Framing of Information Seeking

One of the most powerful – and subtle – strategies people can use to arrive at desired conclusions is to frame the questions they ask in a biased manner, making confirmation of a desired conclusion more likely than disconfirmation. Imagine that a person wants to assess their social abilities. They could begin by asking themselves *Am I extroverted?* or they could begin by asking themselves, *Am I shy?* These two questions may seem like two sides of the same self-assessment coin, but a long-standing body of work shows that people answer these two questions in very different ways: Ask people if they are extroverted and they will primarily search for positive evidence that they are extroverted. After finding it, people will tend to believe they are outgoing, gregarious individuals. However, ask the same people if they are shy, and they will tend to search for evidence that they are reticent and private people, which can lead them to think that they are more reserved and inhibited (Kunda, Fong, Sanitioso, & Reber, 1993; Snyder & Swann, 1978).

A number of factors can influence the question people initially pose to themselves and thus the conclusion they ultimately reach. The frame of a question can be suggested by an outside source, such as a salesman or therapist, or from a recent experience. Our assertion, though, is that a person's directional motives may possess a similar influence. It is likely that most people favor asking questions that suggest positive conclusions over negative ones. They would prefer to ask themselves if they are intelligent rather than stupid, good-looking rather than unattractive, healthy rather than unhealthy, and, outgoing rather than shy. Once those question frames are in place, a confirmatory bias takes over that favors favorable conclusions over alternatives (Pyszczynski & Greenberg, 1987; Mussweiler & Strack, 1999). Indeed, Mussweiler (2003) has argued that such a confirmatory process underlies any number of social comparisons that people engage in to uphold their positive self-views.

Motives Influence the Quantity and Quality of Information Sought

But, sometimes, even if people adopt a congenial question frame, the evidence they gather may not conform to their wishes. Motives likely influence what people do when they encounter favorable versus unfavorable information, altering how much evidence and what quality of evidence people demand before they can cut off an information search and move toward a conclusion.

People show a tendency not to need much evidence in favor of conclusions they like. However, when it comes to conclusions they would rather avoid, they show a marked tendency to demand more evidence and to place whatever uncongenial evidence they have under intense scrutiny. In a phrase, people seem to adopt a “can I believe” stance to favorable conclusions: Any evidence permitting them to believe favorable conclusions is taken at face value. For unfavorable conclusions, they adopt a “must I believe” stance, asking instead if they are compelled to believe an unfriendly message (Gilovich, 1991; Dawson, Gilovich, & Regan, 2002).

The different treatment given to favorable versus unfavorable evidence is best illustrated by work on *motivated skepticism* (Ditto & Lopez, 1992), which has shown that people accept congenial information more or less effortlessly, but treat uncongenial information with a fine-toothed comb. In one study, participants were given a kit to test for the (actually fictitious) medical condition TAA deficiency, which they were told was linked to unfortunate pancreatic disorders later in life. Participants were instructed to spit into a cup, dip a testing strip into the saliva, and wait for it either to change color or stay the same (to indicate the absence or presence of the deficiency).

Ditto and Lopez (1992) covertly observed that participants receiving “good news” – the test result indicating no increased risk for pancreatic problems –accepted the information readily. They waited less time for a color change in the strip and they did not engage in any re-testing behaviors (re-dipping the strip, for example, just to be sure of the result). Those who received the “bad news,” on the other hand, waited longer before accepting the results of the test and engaged in re-testing behaviors three times more often than their peers.

Skepticism was also observed in participants' explicit responses to the test. Those who received the “bad news” thought the test was less accurate and the conclusions less severe, than did those who received the “good news.” In follow-up work, participants proved ready to discount an unfavorable test if there was a reason it might be invalid, but did not discount a favorable test if it contained the same flaw (Ditto, Scepansky, Munro, Apanovitch, & Lockhart, 1998).

Motives Guide the Interpretation of Key Social Concepts and Behaviors

Consider a trait that most of us would like to claim: gifted scholarship. Of course, not every academic can rightfully claim the trait, but many do, probably more than are justified. What allows them to make the claim? One possibility stems from the fact that the trait itself is ambiguous. What exactly constitutes *gifted scholarship*? Is it the number of publications one has? Professional accolades? Or perhaps it is a trait revealed by teaching evaluations. In essence, one way people may achieve motivated conclusions about themselves involves the way they define the traits and abilities that they wish to possess.

Self-Serving Definitions of Social Concepts

Each scholar probably possesses her own idiosyncratic standard or definition of good scholarship, and, for most people, that idiosyncratic definition is probably self-serving in nature. A great deal of research (for reviews, see Critcher, Helzer, & Dunning, 2010; Dunning, 1999) suggests that the definitions people assign to positive traits (e.g. *sophisticated*) just happen to be the very qualities that they, themselves, possess (Dunning & Cohen, 1992; Dunning, Meyerowitz, & Holzberg, 1989; Dunning, Perie, & Story, 1991; Dunning & McElwee, 1995). A wine expert emphasizes knowledge of fine wines in her definition of "sophisticated." A person who has read many books emphasizes a more bookworm variant of the term. In contrast, the qualities people assign to unfavorable traits (e.g., *submissive*) tend to be qualities they fail to see in themselves. Thus, one way people can support desired self-impressions is to form their definition of a particular trait with reference to their own behaviors and qualities.

Note that the key to motivated self-assessment in these examples is trait ambiguity--assessing the self on traits that can be defined in any number of ways. As evidence of this, when researchers reign in people's ability to exploit ambiguity in their definitions by assigning them a specific and concrete definition of an otherwise ambiguous trait (e.g., they are presented a definition of *sophisticated* that includes only being able to speak foreign languages and cooking gourmet meals), people offer self-assessments that are significantly less self-enhancing (Dunning et al., 1989).

Self-Serving Labels Applied to Behavior

Another avenue for motivated reasoning in self-understanding lies at the intersection of a concrete instance of behavior and the more abstract label a person assigns to it. Take the act of completing a tedious laboratory task (e.g., writing number words – *one, two, three, four* –

continuously for four minutes): how might a person make meaning out of that behavior? Jordan and Monin (2008) put participants in such a situation and later asked them to assess their own moral qualities and the moral qualities of another participant in the experiment. Under normal circumstances (i.e., when participants were led to believe that both they and the other participant had been made to complete the exact same task), people did not tend to imbue the act of completing the task with particular self-serving meaning. They rated their own morality and the morality of the other participant roughly equally. In another condition, participants simply witnessed the other participant (actually a confederate) refuse to do the task before rating the other participants' morality. Here, again, participants assessed themselves and the confederate roughly equally.

However, in the key condition, participants were put in a position that was potentially threatening to their sense of self. Having just spent four minutes completing the dreary and tedious task, participants witnessed a confederate refuse to complete the very same task with no adverse consequence. Faced with the uncongenial possibility that they had just played the sucker, participants showed a motivated shift in thought. Participants rated their own morality (but not their intelligence, confidence, or sense of humor) as higher than the morality of the participant who refused to complete the task. In participants' minds, their efforts at the banal task were demonstrative of moral character rather than gullibility. A second study confirmed the motivational nature of this shift in moral self-perceptions, in that it did not arise if participants first engaged in a self-esteem building exercise.

Keeping Accounts of Behavior

A final way people maintain favorable views of self is to balance the choices they make against one another to e~~i~~nsure that, on balance, they send adequate signals that they are worthy and capable individuals. That is, people at times can indulge in less-than-admirable behavior if they have first built up their “bona fides” as moral and respectable individuals. By building up accounts of desirable behavior, they obscure any chance that their indulgent transgressions can be interpreted as anything significant about their overall character.

Work on *moral licensing* shows this balancing most directly, and thus reveals how people can engage in questionably immoral or stigmatized behavior without worrying about the dishonor often associated with that behavior, provided that they feel that have adequately signaled their moral worth elsewhere. For example, when people are feeling particularly moral, they may--paradoxically--give themselves permission to act in morally questionable ways with no consequence for their own self-views. In a well-known demonstration, Monin and Miller (2001) offered participants an opportunity to signal their good moral credentials by disagreeing with a number of sexist views, and then observed their responses to a subsequent task in which they chose the best candidate for a job usually linked to men. Relative to control participants, participants who had just established their non-sexist credentials by rejecting blatantly biased statements made *more* biased hiring decisions, in that they were more likely to pick a man over a woman for a stereotypically male job in law enforcement.

In similar set of studies, Sachdeva, Iliev, and Medin (2009) showed that people were *less* charitable than control participants following a reminder of their own positive traits and *more* charitable than controls following a reminder of their own negative traits. Such a result makes little sense unless one posits the idea that people engage in moral accounting to ensure that they can see themselves as good and moral agents. Thus, if one has already been given the opportunity to signal their moral worth (via a recitation of their positive qualities), they need not repeat costly moral behavior and can instead engage in some questionable actions. However, when this signal has been interrupted or undermined by a reminder of one's shortcomings, people are energized to restore a positive self-image via moral behavior.

Current Issues in Motivated Reasoning Theory and Research

Decades of research on motivated reasoning has offered a number of classic demonstrations of its operation in people's day-to-day thinking about themselves. Without a doubt, motivated reasoning exists and serves as a primary line of defense that protects people's self-views from potentially damaging or conflicting information. But the exact implications of motivated reasoning for social thought and life remain somewhat controversial, spurring current research and enduring debates.

Do Congenial Conclusions Necessarily Reflect Motivated Reasoning?

One issue concerns the frequency and scope of motivated reasoning. People come to congenial conclusions all the time, but that does not necessarily mean that those conclusions were inspired by motivated reasoning. A number of flaws in thinking, including everyday heuristics for evaluating information may give rise to congenial conclusions even without any particular biasing motivation.

Consider one of the most common and visible biases suggesting motivated reasoning in self-perception, the above-average effect, the statistically impossible phenomenon in which the average person rates his or her skill, on average, as way above average (Alicke, 1985; Alicke & Goverum, 2005; Dunning, Heath, & Suls, 2004). Although the above-average effect looks like it must be a product of motivated reasoning, this is not necessarily the case. Other nonmotivational quirks of thinking can lead to a good deal of above-average responding.

For example, one could propose that the above-average effect arises because self-serving motivations prompt people to ask questions in such a way that almost guarantees a favorable self-impression (e.g., they ask whether they are more intelligent than their peers rather than whether their peers are more intelligent than them). Such question frames, however, can just as easily be prompted by nonmotivational factors, producing above-average effects in contexts that clearly have nothing to do with motivation. Consider three workshop tools that one might own: a hammer, a screwdriver, and a saw. If someone were to ask you if the hammer is a more useful tool than the other two, you are likely to agree (i.e., it is good for pounding nails, and opening nuts when you cannot find the nutcracker)—making the hammer an above-average tool. But here is the trick. You are likely to do the same for the screwdriver if asked about it (e.g., it’s essential for screws, and for opening balky cans) and for the saw (e.g., it cuts wood cleanly, and in an emergency can be used as a musical instrument).

That is, if one is given a pointed hypothesis by an external agent about one object within the group, one is likely to hunt for information that confirms the hypothesis. Having found confirmatory evidence, one will conclude (perhaps wrongly, but not because of any particular motivation) that the object in question is above average. But, here's the rub: Because each tool is useful in its own way, and the ways in which a particular tool is more useful than the others become salient when that tool is the focus of attention, it is easy to slip into claiming that the tool is more useful than others (Giladi & Klar, 2002; Klar & Giladi, 1997). That is, even though the guiding question is a comparative one, people often fail to complete the comparison in their thinking, focusing on the central object and thinking about how useful that particular object is (Weinstein & Lachendro, 1982; Kruger, 1999).

Does this mean that the above average effect is a really non-motivated phenomenon? Not necessarily, since current research suggests that motivational dynamics are often in play in producing the effect. In one illustrative study, participants were asked to rate themselves along 23 different trait dimensions. Four to eight weeks later, they were handed the exact same set of ratings and told that they came either from themselves or from some other student at their university. They were then asked to provide the rating that a “typical student” in their university would give. When participants thought that the ratings came from someone else, their ratings for the “typical student” closely mimicked the ratings they were given. However, when participants were reminded that the ratings were actually their own, their assessments of the “typical student” changed, with participants rating the student much more negatively than they did in the other group. Motivated to think of themselves as superior to others, participants now seemed impelled to maintain a gap in perception between self and other that favored the self. Such a motivation was absent when they considered ratings that they thought came from another person (Guenther & Alicke, 2010).

Does Motivated Reasoning Necessarily Lead to Error in Self-Knowledge?

The question of whether motivated reasoning leads to error in self-knowledge is also a topic of disagreement. On first pass, it may seem almost impossible that the sorts of cognitive tricks prompted by motivation would fail to lead to some kind of systematic distortion in self-understanding. Surely, differential treatment of flattering versus unflattering evidence should lead to errors and omissions in a person's global self-evaluation.

Self-Immunization

An argument can be made, however, that people can hold vaulted views of themselves yet still maintain an accurate impression of the world and one's place in it. Theorizing by Greve and Wentura (2003) suggests that people can achieve this happy state by simultaneously holding uniformly positive self-evaluations at an abstract (i.e., trait) level while knowing their limitations and weaknesses at a more concrete (i.e., behavioral) level. Greve and Wentura have labeled this possibility *self-immunization*.

As one demonstration of accuracy at the concrete level but self-enhancement at the abstract level, Greve and Wentura (2003) asked participants to tackle a general knowledge test in which they were asked a variety of questions from four different domains of knowledge (e.g., politics, science, history, art). They competed against a confederate, allowing the experimenters to randomly assign participants to perform better than the confederate on two of the four domains and worse on the other two. Later, when participants were asked to offer self-assessments, Greve and Wentura found that participants were able to accurately recall which areas of knowledge they excelled at and which areas they failed at – that is, they showed no evidence of distorting the raw data of their concrete experience in a self-serving direction. At the behavioral level, they were largely accurate.

However, at a more abstract or trait level, participants showed no allergy to self-enhancement. When asked which domains were most indicative of one's overall intelligence, participants tended to choose the domains in which they themselves had purportedly beaten the confederate. Thus, at a more "conceptual" level, participants could exploit ambiguity to think well of themselves, even though they remained accurate at a more concrete level. Studies like these (for a demonstration of self-immunization using implicit measures, see Wentura & Greve, 2004) suggest that people may be able to maintain knowledge of their concrete strengths and weaknesses, but may put a spin on this information as a way of maintaining positive global self-views (for a similar view, see Armor & Taylor, 1998).

In effect, Greve and Wentura argue that people can reasonably possess rosy self-impressions yet maintain accuracy, too. Note, though, that to maintain this ideal balance, people must stay aware of the often hazy but critically important line between their *knowledge*, i.e., information about concrete behavioral performance, and *inference*, i.e., the more flattering opinions they extract about themselves at the abstract trait level (see Critcher, et al., 2010; Schneider, 2001). To the extent that they are successful at recognizing where actual information ends and more subjective opinion or “spin” begins, they can have both positive impressions about the self yet maintain accurate self-insight. For example, Erik can believe in his subjective inference that he is more “sports savvy” than David because he knows (objectively) more about baseball. And David can be just as justified in extolling his own subjective evaluation that his sports savvy is excellent because he knows (objectively) more about soccer than Erik does.

However, if they wish to remain accurate in their self-assessments and not just enhancing, Erik must acknowledge that his “sports savvy” is primarily in the domain of baseball; David must acknowledge that his “sports savvy” is primarily in soccer; and both must acknowledge that their “sport savvy” likely would not necessarily translate to expertise in football, basketball, tennis, golf, cricket, polo, rugby, wrestling, boxing, and virtually every other sport. Without this understanding, the happy marriage between (abstract) self-enhancement and (concrete) self-knowledge begins to dissolve.

Put another way, as the line between subjective interpretation and objective knowledge blurs, the potential for accuracy in the realm of self-insight diminishes. In the example above, if Erik or David infers from his self-perceived sports savvy (rather than a more concrete review of his specific knowledge of baseball or soccer) that he can beat the other in a sport trivia contest, one of them is by definition going to be in error. Or, to the extent that their subjective inferences cause them to distort either on-line or retrospective accounts of their objective performance, they will again be led to distorted self-views.

Indeed, much empirical work suggests that people have a very difficult time keeping their subjective inference and objective knowledge straight. First, people tend not to think of their inferences as subjective. They create self-serving templates of intelligence, leadership, and so on, and construe them as universal—just as applicable to others as they are to the self (for reviews, see Dunning, 1999; 2002a, 2002b). For example, Erik will use expertise in baseball as a cue to expertise about sports in general; David will do the same with his expertise in soccer. Second, people's subjective self-views prompt them to misremember objective performances. People with high impressions of themselves tend to believe they have performed better in the past than they have actually done. People with low impressions tend to underestimate how well they have done (McFarland, Ross, & DeCourville, 1989; Story, 1998).

In addition, people's abstract notions of themselves also influence how well they think they are achieving on any concrete task as they complete it. Those who believe they are skilled think they are doing better than those who think they are less skilled, even when equating for actual performance. In one demonstration of this phenomenon, Ehrlinger and Dunning (2003, Study 1) asked participants to evaluate their abstract reasoning ability before completing a standardized test of that sort of reasoning. Once they had completed the test, participants evaluated how they had performed relative to their peers on the test. Noting that participants showed the usual self-enhancement bias on their performance estimates (on average, people thought they performed in the 61st percentile), Ehrlinger and Dunning examined the relationship between participants' chronic self-views (measured before the task), their actual performance, and their estimated performance. The results indicated that participants' estimates of their objective performance were predicted by their subjective self-views, but not by their actual performance.

That is, participants based their performance evaluations on their *a priori* beliefs about their abilities, and showed little sensitivity to actual performance. To be sure, participants' self-views were not totally divorced from reality (the one-item measure of people's global self-perceptions of ability significantly predicted actual performance), and, as such, stable self-views were not useless bases for self-evaluation. However, the broader point is that participants leaned heavily (indeed, too heavily) on these subjective beliefs when predicting a single instance of performance and showed little insight into their concrete performance on the task at hand (see also Critcher & Dunning, 2009).

Looking across the prediction literature, there are numerous examples of people making poor concrete self-predictions about the future because they base those predictions on subjective self-views (Buehler, Griffin, & Ross, 1994; Epley & Dunning, 2000; Koehler & Poon, 2006). As one striking example, medical students' self-rated ability at the end of medical school correlates strongly ($r > .50$) with their self-related ability in the first year of medical school, even though the former ratings tend to be unrelated to more objective measures of performance ability, including supervisor ratings and final board exams (Arnold, Willoughby, & Caulkins, 1985). Taken together, findings like these raise serious doubts about whether people can distinguish between concrete information about themselves from more abstract (and flattering) self-evaluations. Thus, to our minds, inaccuracies at the abstract level of self-evaluation pose serious threats to self-knowledge, broadly construed. However, Greve and Wentura (2003) make an opposite claim, with other theorists more in the middle (e.g., Schneider, 2001), meaning that more analysis and empirical study on this issue would be worthwhile.

Can Motivated Reasoning Be Corrected?

If motivated reasoning leads to error, how might one overcome it? There are two techniques that have been proposed, one straightforward and the other trickier. The straightforward way to correct for motivated reasoning applies to many other reasoning problems as well. Because people tend to look for confirming information, and often stop short of considering all the information they should, one simple way to correct for biases—including motivated ones—is to explicitly *consider the opposite.* For example, if one's thoughts are leading one to conclude, with confidence, that one will obtain a well-paying job after college, or a seat at a highly-ranked law or medical school, the best corrective is to explicitly consider why such a conclusion might be wrong. What could go wrong to prevent that job or post-graduate career? What could one have forgotten to do to make those outcomes more assured?

Much work shows that considering the opposite does a good deal to remove biases in people's conclusions, especially those promoted by confirmatory thinking (Lord, Lepper, & Preston, 1984), anchoring (Mussweiler, Strack, & Pfeiffer, 2000), and the hindsight bias (Arkes, Faust, Guilmette, & Hart, 1988). This technique works even better than simple exhortations to be fair and impartial in decision-making. Explicitly asking people why a prediction might be wrong causes them to be significantly less overconfident in that prediction (Hoch, 1985; Koriat, Lichtenstein, & Fischhoff, 1980).

A second technique works specifically to alleviate motivated biases, and involves asking people to engage in *self-affirmation*. In self-affirmation, individuals consider an aspect or part of themselves that they hold in high regard and which makes them proud (e.g., friends and family, scientific values, artistic values, etc.). They write a short essay about a time they were proud of something related to that self-aspect. The net effect of this exercise is that people are subsequently much more likely to accept threatening information (for reviews, Sherman & Cohen, 2006; Steele, 1988). For example, doing a self-affirmation exercise makes people more likely to accede to the idea that they are at risk for HIV, and to purchase more condoms as a response (Sherman, Nelson, & Steele, 2000). Such exercises make people more open-minded; that is, more willing to listen to and consider the arguments of people who disagree with their own positions on politics and morality (Cohen, Aronson, & Steele, 2000). To be sure, no one knows exactly why self-affirmation exercises work—they just do, as has been demonstrated in a wide variety of domains of some consequence to the people involved (see Sherman & Cohen, 2006).

Contemporary and Future Questions

Questions about motivated reasoning still exist, and are at the center of current (and, we believe, future) empirical research in psychology and related fields. Some of these questions are classic and enduring ones for which methodological and theoretical advances now make them amenable to empirical study. Let us consider three such questions in turn.

The Relation of Motivated Reasoning to Self-Deception

The first question focuses on the relationship between motivated reasoning and self-deception. Traditionally, the two concepts have not been considered the same thing (see Paulhus, this volume, for a discussion of self-deception). People can surely reason their way toward beliefs that are not true, but that is not exactly what is meant by self-deception. As traditionally defined, self-deception involves the paradoxical situation in which a person believes *X* to be true, but convinces themselves of not-*X*. Philosophers have long argued about how, or whether, such a bifurcated belief system could be maintained (see Mele, 1997, 2001).

However, if one relinquishes this strict definition, motivated reasoning emerges as a paradigmatic case of self-deception. Suppose that the quest for self-knowledge involves an inferential race between a correct belief about one's self ("When it comes to relationships, I am only about as caring as the average person") and an incorrect, but flattering one ("I'm probably a better boyfriend than most guys I know."). Now, suppose that a person's capacity for motivated reasoning allows the person to construct and alter the race course so that one belief (the congenial, but incorrect one) will almost always win out over the other (Mele, 1997)— and will do so without leaving the slightest trace of suspicious play. An interesting question for future research is how people alter the race course without awareness of the effort to do so. We can offer two possibilities for future empirical work.

Motivated Reasoning Can Operate Nonconsciously

One way for motivated reasoning to do its work without leaving a conscious trace is to situate its operations outside of awareness. Many cognitive tasks are completed nonconsciously. Just to speak a sentence, a person needs to choose words, conjugate verbs, and rearrange words into comprehensible phrases, clauses, and sentences—and much of this work takes place before any product of it reaches consciousness (Bargh, 1994).

Recent work suggests that the impact of motivated reasoning extends down into processes that are clearly nonconscious. For example, what the visual system presents to consciousness can be shaped by motivated preferences. In one illustrative study, Balcetis and Dunning (2006) told participants that a computer was about to assign them to one of two fates. The first was pleasant, and involved drinking some freshly-squeezed orange juice. The second was unpleasant, and involved drinking a foul-smelling pink and green concoction euphemistically described as an "organic garden smoothie." Some participants would be assigned to the orange juice if the computer showed them a letter of the alphabet; for the others, the computer had to show them a number. The computer then showed them an image that looked like it could be either the letter "B" or the number "13." Participants tended to report seeing the image that assigned them to the orange juice much more often than they did the image that assigned them to the smoothie. Subsequent experiments demonstrated that participants truly did see the image they wanted to see and were not merely lying to the experimenter.

Motivated Reasoning Resides in the Past

Another trick about motivated reasoning is that it need not always be actively motivated. That is, once people have formed a congenial belief (e.g., I am an excellent driver/student/cook), that crystallized belief is available for them for the foreseeable future. It merely becomes part of their belief system, and need not be "re-distorted" by motivated reasoning. Thus, one can often reach distorted conclusions based on illusory self-beliefs that have been crafted long and ago thus need no additional motivated reasoning in the present. In this way, the motivated and distorted nature of these beliefs remains hidden, and the person stays blissfully self-deceived.

To date, there is no research we are aware of that shows that motivated beliefs, once crystallized, remain in the person’s belief system to distort future conclusions about the self. However, one recent series of studies has shown evidence of at least the first step in the process. Motivated beliefs, once formed, tend to stick—that is, to become functionally autonomous from the circumstances that created them. In its specifics, this work asked whether the timing of self-affirmation exercises mattered (Critcher, Dunning, & Armor, 2010). Would self-affirmation prevent people from acting defensively toward threatening information if the affirmation came after the threat rather than before?

To address this question, Critcher et al. (2010) asked some participants to complete self-affirmation exercises before they responded to a threat—such as failing a test. Others completed the self-affirmation only after the threat had been responded to. The researchers found that self-affirmation before the threat tended to stop people from being defensive, the usual self-affirmation result. However, if the self-affirmation came after people had already responded to the threat through defensive self-enhancement, the exercise did nothing to reduce the amount of defensive resistance to the threat that people displayed. That defensive reaction, now completed, had crystallized and was not “undone” by the introduction of self-affirmation. That is, once in place, those defensive conclusions were presumably positioned to distort future conclusions that the individual might reach about themselves and their place in the world.

Relation of Motivated Reasoning to Reality Constraints

In her ground-breaking review article on motivated reasoning, Kunda (1990) asserted that people maintain a nuanced dance between their wishes on the one hand and reality on the other (echoing Heider, 1958, who offered a similar analysis of people's need to balance desired conclusions with conclusions warranted by data). People do tend to reach favorable conclusions, but those conclusions are heavily hemmed in by "reality constraints." Indeed, work in the 1990s showed just how much reality constraints matter. People tend to provide unrealistically positive self-views, for example, only when the traits in question are ambiguous enough for motivated reasoning to have some leeway to provide favorable interpretations. Instead, when the meaning of a trait was clear (e.g., *punctual, neat, mathematically skilled*), researchers saw *no* distortion in self-judgment, despite obvious desires to see oneself as positively as possible (Dunning et al., 1989). Hsee (1995, 1996) also showed that motivated reasoning produced distorted judgments only when the factors justifying conclusions were elastic in their application—that is, there was some play in what factors were relevant versus not. When those factors were more clear-cut in their applicability, no motivated distortion was found.

Recent current events, however, suggest that reality constraints may not be as restrictive as this past work suggests. As of the writing of this chapter, people can be shown to believe ideas that have been patently disproved time and again to be false. For example, despite voluminous evidence to the contrary, in February 2011 a full 51% of respondents planning to vote in the 2012 Republican Presidential primaries believed that the current U.S. President, Barack Obama, was not born in the country, and thus ineligible for his office, with another 28% unsure (Public Policy Polling, 2011). Of Americans, roughly 20% thought that Obama was Muslim, with such perceptions more widespread (34%) among those (conservative Republicans) who presumably oppose his presidency (Pew Forum on Religion & Public Life, 2010).

Thus, it appears that a re-examination of the interplay between motivation and reality constraints may be in order. When are motivated distortions reigned in by factual evidence? And when does motivation triumph despite evidence and constrictions from the real world?

Perception of Motivated Reasoning in Self versus Other

The final question has to do with how well people understand the impact of motivated reasoning in their everyday world and life. To be sure, one question that seems settled is that people underestimate just how much their own judgments are molded by wishes, preferences, and fears. In study after study, people describe themselves as more unbiased than their peers (Pronin, Lin, & Ross, 2002; Pronin & Kugler, 2007), whether the bias involved be motivational or not.

But how calibrated are people in their beliefs about the extent to which their peers' thoughts are guided by motivated wishes and desires? Do people understand, for example, how much other people rationalize? Or do they over- or underestimate it? Given the commonness of motivated reasoning in everyday life, getting its impact right would seem to be an important source of social wisdom.

To date, there are very few investigations of people's understanding of motivated reasoning in the social world. One extant study, for example, finds that people overestimate motivated reasoning—that people are "naïve cynics." People expect that their peers will take too much credit for positive contributions to joint projects, but will deny responsibility for setbacks. It turns out that although that assumption is correct for the former, it is not for the latter. People are just as likely to accept their share of blame for failures—even overdoing it—as they are to take credit for successes (Kruger & Gilovich, 1999). Further work, however, will be needed to flesh out whether this is a general tendency or just an isolated instance of a social fallacy.

Concluding Remarks

To understand self-understanding, one must study closely the impact of motivations on people’s reasoning about the self and their social world. To be sure, people want to achieve an accurate understanding of themselves, but they seem not to mind holding flattering impressions of themselves, as well. Thus, to gauge how people come to understand themselves—and, more importantly, how they *mis*understand themselves, one must first grasp the when, how, and why of motivated reasoning. As researchers interested in people's capacity for genuine self-knowledge, we must take seriously not only the pervasiveness of motivated processes in self-understanding, but also the nuances that govern the interplay between self-flattery and self-insight. We must further understand that motivated processes present difficult challenges to the Delphic admonition for people to “know thyself.”

References

Abelson, R. P., Aronson, E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J., & Tannenbaum R. H. (1968). *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally.

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49,* 1621-1630.

Alicke, M.D., & Govorun, O. (2005). The better-than-average effect. In M.D. Alicke, D.A. Dunning, & J. I. Krueger, (Eds.*), The self in social judgment* (pp. 85-106). New York: Psychology Press.

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

Armor, D.A., & Taylor, S.E. (1998). Situated optimism: Specific outcome expectancies and self-regulation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 309-379). New York: Academic Press.

Arnold, L., Willoughby, T. L., & Calkins, E. V. (1985). Self-evaluation in undergraduate medical education: A longitudinal perspective. *Journal of Medical Education, 60*, 21-28.

Balcetis, E., & Dunning, D. (2006). See what you want to see: The impact of motivational states on visual perception. *Journal of Personality and Social Psychology, 91*, 612-625.

Bargh, J. A. (1994). The Four Horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1-40). Hillsdale, NJ: Erlbaum.

Baumeister, R. F., & Newman, L. S. (1994). Self-regulation of cognitive inference and decision processes. Personality and Social Psychology Bulletin. 20, 3-19.

Beckman, L. (1973). Teachers' and observers' perceptions of causality for a child's performance. *Journal of Educational Psychology, 65*, 198-204.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology, 67*, 366-381.

Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin, 26*, 1151-1164.

Critcher, C. R., & Dunning, D. (2009). How chronic self-views influence (and mislead) self-assessments of performance: Self-views shape bottom-up experiences with the task. *Journal of Personality and Social Psychology, 97*, 931-945.

Critcher, C. R., Dunning, D., & Armor, D. A. (2010). When self-affirmation reduces defensiveness: Timing is key. *Personality and Social Psychology Bulletin, 36*, 947-959.

Critcher, C. R., Helzer, E. G., & Dunning, D. (2010). Self-enhancement via redefinition: Defining social concepts to ensure positive views of self. In M. D. Alicke & C. Sedikides (Eds.), *Handbook of self-enhancement and self-protection* (pp. 69-91). New York: Guilford Press.

Dawson, E., Gilovich, T., & Regan, D. T. (2002). Motivated reasoning and performance on the Wason selection Task. *Personality and Social Psychology Bulletin, 28*, 1379-1387.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology, 63*, 568-584.

Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovitch, A. M., & Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology, 75*, 53-69.

Dunning, D. (1999). A newer look: Motivated social cognition and the schematic representation of social concepts. *Psychological Inquiry, 10,* 1-11.

Dunning, D. (2001). On the motives underlying social cognition. In N. Schwarz & A. Tesser (Eds.) *Blackwell handbook of social psychology: Volume 1: Intraindividual processes* (pp. 348-374). New York: Blackwell.

Dunning, D. (2002) a. The relation of self to social perception. In M. Leary and J. Tangney (Eds.), *Handbook of Self and Identity* (pp. 421-441). New York: Guilford.

Dunning, D. (2002) b. The zealous self-affirmer: How and why the self lurks so pervasively behind social judgment. In S. Fein & S. Spencer (Eds.) *Motivated social perception: The Ontario symposium* (vol. 9, pp. 45-72), Mahwah, NJ: Erlbaum.

Dunning, D. (2007). Self-image motives and consumer behavior: How sacrosanct self-beliefs sway preferences in the marketplace. *Journal of Consumer Psychology, 17,* 237-249.

Dunning, D., & Cohen, G. L. (1992). Egocentric definitions of traits and abilities in social judgment. *Journal of Personality and Social Psychology, 63*, 341-355.

Dunning, D., Heath, C., & Suls, J. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5,* 69-106.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57,* 1082-1090.

Dunning, D., & McElwee, R. O. (1995). Idiosyncratic trait definitions: Implications for self-description and social judgment. *Journal of Personality and Social Psychology, 68*, 936-946.

Dunning, D., Perie, M., & Story, A. L. (1991). Self-serving prototypes of social categories. *Journal of Personality and Social Psychology, 61*, 957-968.

Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology, 84*, 5-17.

Epley, N., & Dunning, D. (2000). Feeling "holier than thou": Are self-serving assessments produced by errors in self or social prediction? *Journal of Personality and Social Psychology, 79*, 861-875.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Giladi, E. E., & Klar, Y. (2002). When standards are wide of the mark: Nonselective superiority and inferiority biases in comparative judgments of objects and concepts. *Journal of Experimental Psychology: General, 131*, 538-551.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*. New York: Free Press.

Greve, W., & Wentura, D. (2003). Immunizing the self: Self-concept stabilization through reality-adaptive self-definitions. *Personality and Social Psychology Bulletin, 29*, 39-50.

Guenther, C. L., & Alicke, M. D. (2010). Deconstructing the better-than-average effect. *Journal of Personality and Social Psychology, 99*, 755-770.

Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology, 21*, 107-112.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 719-731.

Hsee, C. K. (1995). Elastic justification: How tempting but task-irrelevant factors influence decisions. *Organizational Behavioral and Human Decision Process, 62*, 330-337.

Hsee, C. K. (1996). Elastic justification: How unjustifiable factors influence judgments. *Organizational Behavior and Human Decision Processes, 66*, 122-129.

Jordan, A. H., & Monin, B. (2008). From sucker to saint: Moralization in response to self-threat. *Psychological Science, 19*, 809-815.

Klar, Y., & Giladi, E. E. (1997). No one in my group can be below average: A robust positivity bias in favor of anonymous peers. *Journal of Personality and Social Psychology, 73*, 885-901.

Koehler, D. J., & Poon, C. S. K. (2006). Self-predictions overweight strength of current intentions. *Journal of Experimental Social Psychology, 42*, 517-524.

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

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, 77, 221-232.

Kruger, J., & Gilovich, T. (1999). "Naive cynicism" in everyday theories of responsibility assessment: On biased assumptions of bias. *Journal of Personality and Social Psychology*, 76, 743-753.

Kruglanski, A. W., Dechesne, M., Orehek, E., & Pierro, A. (2009). Three decades of lay epistemics: The why, how and who of knowledge formation. *European Review of Social Psychology, 20*, 146-191.

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology, 53*, 37-54.

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

Kunda, Z., Fong, G. T., Sanitioso, R., & Reber, E. (1993). Directional questions direct self-conceptions. *Journal of Experimental Social Psychology, 29*, 63-86.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, *47,* 1231-1243.

McFarland, C., Ross, M., & De Courville, N. (1989). Women's theories of menstruation and biases in recall of menstrual symptoms. *Journal of Personality and Social Psychology, 57*, 522-531.

Mele, A. R. (1997). Real self-deception. *Behavioral and Brain Sciences, 20,* 91-136.

Mele, A. (2001). *Self-deception unmasked*. Princeton: Princeton University Press.

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology, 81*, 33-43.

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110*, 472-489.

Mussweiler, T., & Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology, 35*, 136-164.

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

Psychology Bulletin, *26*, 1142-1150.

Pew Forum on Religion & Public Life. (2010). *Growing number of Americans say Obama is a Muslim*. Unpublished manuscript, Pew Forum on Religion & Public Life. Washington, DC.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin, 28,* 369-381.

Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology, 43*, 565-578.

Public Policy Polling. (2011). *Huckabee tops GOP field, 51% are birthers and love Palin.* Unpublished manuscript, Public Policy Polling, Raleigh, North Carolina.

Pyszczynski, T., Greenberg, J., & Holt, K. (1985). Maintaining consistency between self-serving beliefs and available data: A bias in information evaluation following success and failure. *Personality and Social Psychology Bulletin, 11*, 179-190.

Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. *Advances in Experimental Social Psychology, Vol. 20* (pp. 297- 340). Elsevier: New York.

Redlawsk, D. P., Civettini, A. J. W., & Emmerson, K. M. (2010). The affective tipping point: Do motivated reasoners ever "get it"? *Political Psychology, 31*, 563-593.

Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science, 20*, 523-528.

Schneider, S. L. (2001). In search of realistic optimism: Meaning, knowledge, and warm fuzziness. *American Psychologist, 56,* 250-263.

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. In M. P. Zanna (Ed.) *Advances in Experimental Social Psychology* (Vol. 38, pp. 183-242). San Diego, CA: Academic Press.

Sherman, D. K., Nelson, L. D., & Steele, C. M. (2000). Do messages about health risks threaten the self? Increasing the acceptance of threatening health messages via self-affirmation. *Personality and Social Psychology Bulletin, 26,* 1046-1058.

Snyder, M., & Swann, W.B. (1978). Hypothesis-testing in social interaction. *Journal of Personality and Social Psychology, 36,* 1202-1212.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (vol. 21, pp. 261-302), San Diego: Academic Press.

Story, A. L. (1998). Self-esteem and memory for favorable and unfavorable personality feedback. *Personality and Social Psychology Bulletin, 24*, 51-64.

Taylor, S. E. (1983). Adjustment to threatening events: A theory of cognitive adaptation. *American Psychologist, 38*, 1161-1173.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103,* 193-210.

Weinstein, N. D., & Lachendro, E. (1982). Egocentrism as a source of unrealistic optimism. *Personality and Social Psychology Bulletin, 8,* 195-200.

Wentura, D., & Greve, W. (2004). Who wants to be…erudite? Everyone! Evidence for automatic adaptation of trait definitions. *Social Cognition, 22*, 30-53.

Willis, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin, 90,* 245-271.

Wyer, R. S., & Frey, D. (1983). The effects of feedback about self and others on the recall and judgments of feedback-relevant information. *Journal of Experimental Social Psychology*, *19*, 540-559.