4

Shallow Thoughts about the Self
The Automatic Components of Self-Assessment

THOMAS GILOVICH
Cornell University
NICHOLAS EPLEY
University of Chicago
KARLENE HANKO
Cornell University

It is not our intent to coin a new term, but any review of the pertinent social psychological literature leads to the conclusion that people are prone to an illusion of personal strength. That is, people's assessments of their own abilities to meet various challenges exceed the best dispassionate analyses of those abilities. People read about Milgram's obedience experiments and come away convinced that they, unlike the majority of actual participants in those studies, would be strong enough to stand their ground and disobey the experimenter (Bierbrauer, 1979). People read about the various bystander (non)intervention studies and likewise remain convinced that they would have sufficient strength to overcome the fear of embarrassment and come to the rescue. And people's assessments of their own traits and abilities have been shown, time and time again, to be overly optimistic (see Alicke & Govorun, this volume).

Our aim in this chapter is to shed light on why people are prone to such an illusion of personal strength. This aim is likely to make some readers wonder whether we are prone to the illusion of personal strength ourselves. After all, there are already perfectly satisfactory explanations of all of the manifestations of this illusion. Do we really have anything useful to add? Is another perspective likely to advance our discipline's understanding of these phenomena? Does the discipline really need yet another explanation of the above average effect?

We believe there is still much to be learned about the processes that give rise to the various manifestations of the illusion of personal strength. In particular,

we believe that recent theoretical developments concerning the automaticity of everyday behavior (Ferguson & Bargh, 2004) and the "two systems" that people bring to bear in making judgments (Chaiken & Trope, 1999; Epstein, 1994; Kahneman & Frederick, 2002; Sloman, 1996) can shed new light on overly-optimistic self-assessments. One system of judgment, sometimes called "System 1," is associationist and produces rapid, effortless assessments in parallel fashion. The other, "System 2," is rule-based and deductive, and produces more effortful, deliberative assessments in serial fashion. Our aim in this chapter is to explore how recent work on automaticity and dual-process approaches to judgment might provide a more complete understanding of various forms of personal overconfidence such as the above average effect.

The guiding insight is that such optimistic self-assessments typically result from, or at least begin with, a minimum of reflection and the shallowest of thinking. Unless there are unusually powerful incentives for deeper analysis (Tetlock, 1992), individuals will typically offer "snap" self-assessments that arise automatically from a set of rudimentary mental processes involved in understanding the particular self-assessment context at hand. The question, therefore, is why such System-1-based snap assessments tend to be optimistic. What is it about this sort of shallow thinking that leads so often to an illusion of personal strength? We devote the bulk of this chapter to precisely that question.

We begin by reviewing evidence consistent with our proposition that people's assessments of their traits and abilities—and how their traits and abilities stack up to others or the demands that must be faced—are of a rapid-fire and rather cursory sort. We then examine the kinds of thoughts that are likely to spring to mind when people ask themselves such questions as "Would I have obeyed the experimenter if I were a participant in Milgram's studies?", "Would I become as loony as Elvis Presley, Michael Jackson, or Kim Jong Il if I were surrounded by obsequious aides and my every whim were indulged?", or "Where do I stand among my peers in terms of my ability to get along with others?" We show how the very nature of these thoughts are likely to lead to the kind of favorable verdicts consistent with an illusion of personal strength.

We also contend that the optimistic biases that result from such thoughts are likely to remain even when the deliberative processes of "System 2" are called upon to inspect and perhaps modify the output of System 1. System 1 gets there first, and so its products are likely to channel subsequent System 2 processing, with the result that any corrections to one's initial assessments are likely to be insufficient. Finally, we conclude by discussing sources of variability in self-assessments and by examining the circumstances under which people's self-assessments are likely to be less positive.

EVIDENCE OF THE SHALLOWNESS OF SELF-ASSESSMENT

Most traits and abilities are inherently comparative. To say that one is a terrible carpenter but a competent writer, for instance, requires a comparison of one's

own efforts at a table saw and computer terminal to the efforts of others. So too with personal traits: a person is tall because others are short, witty because others are dull, or attractive because others are less so. Yet research indicates that most people pay scant attention to others' traits or abilities when assessing their own, making the modal subject in the modal self-assessment experiment look anything but deep and reflective.

Most people, for example, believe that they are happier than their peers because they simply ignore how happy their peers might be and answer the comparative question by consulting only their own absolute happiness (Klar & Giladi, 1999). The answer to the simpler question, "How happy am I?" is substituted for the more challenging, "How happy am I compared to others?" Similarly, people's assessments of whether they are above or below average in the ability to meet various challenges or accomplish certain tasks are based primarily on a simple readout of their own relevant capacities or life histories (Kruger, 1999). Because most people can ride a bicycle without difficulty, most people consider themselves to be above average bicyclists. But because most people would have trouble keeping a unicycle upright, most people think they are below average unicyclists. Owning a home strikes most people (most college respondents at any rate) as likely, and so most think they are more likely than most people to own a home; owning an island strikes most people as something of a pipe dream, and so most think they are less likely than others to own one (Kruger & Burris, 2004).

To be sure, the comparative nature of these sorts of assessments is not always emphasized, making some degree of shallow thinking understandable. However, similar effects have been found in contexts in which the comparison between oneself and others is explicit—and obvious. For example, people in performance competitions act as if shared impediments and facilitators will have a bigger impact on them than on their rivals. Thus, people think that they have a better chance of winning a trivia contest if the questions are easy—even though they know that their opponents will be fed the same easy questions. And people bet more on a round of poker if there are more wild cards in play because the wild cards make better hands easier to obtain—never mind that they make such hands easier for one's opponent to obtain as well (Windschitl, Kruger, & Simms, 2003; see also Moore & Kim, 2003).

Not all self-assessments require an interpersonal comparison of oneself and others; some require a comparison of the costs and benefits of a decision, or the strengths and weaknesses of one's case. The outcome of a romantic relationship, for example, depends not only on its joys and other positive features of the relationship, but also on its sorrows, annoyances, and other negative features. And yet people appear to base their predictions of the health and longevity of their own relationships almost exclusively on its positive elements, all but ignoring the negative (MacDonald & Ross, 1999). It is small wonder, then, that people's forecasts of what's in store for their relationships are so often optimistically biased.

On the basis of results like these, it is hard to argue that most people's assessments of their personal traits, abilities, or futures are the product of deep thought and an exhaustive consideration of the pertinent evidence. Quick, cursory thought

about such matters appears to be the norm. The question, then, is why such cursory thinking tends to produce overly optimistic assessments.

WHY ARE SHALLOW THOUGHTS ABOUT THE SELF SO OFTEN OPTIMISTIC?

The key to understanding why people's quick assessments of their traits and abilities are typically so positively skewed is to recognize that most people regard themselves positively and have good intentions. These two facts steer people's initial thoughts down paths that predispose them toward an illusion of personal strength in the overwhelming majority of domains they might consider.

Implicit Self-Esteem and Optimistic Self-Assessments

Ideas advanced in numerous areas of psychology—from balance theory (Heider, 1958), to connectionism (Rumelhart & McClelland, 1986; Smith, 1996), to work on the Implicit Association Test (Greenwald et al., 2002)—converge on the prediction that when two items in a mental network are each strongly linked to a third, the two will tend to become linked themselves. Thus, people who regard themselves favorably and have a positive stance toward, say, recycling are likely to connect themselves and recycling behavior more strongly than those who either lack self-esteem or don't value such behavior. This initial link—however weak or tentative—then serves to "lean" such individuals toward a positive assessment of their recycling efforts.

When the question is fully engaged—"How do I compare to my peers in terms of my recycling habits?"—most people go beyond this initial association, but the association can nonetheless steer these additional efforts in a particular direction. Because people tend to test hypotheses by trying to confirm them (Klayman & Ha, 1987; Slowiaczek, Klayman, Sherman, & Skov, 1992), they are likely to answer questions about their capacities by searching for evidence that they have them, at the expense of evidence that they do not. Note that this search for confirmatory information need not be effortful or deliberate. People do not need to deliberate and *choose* to search for confirmatory information; they just do. When one considers the question, "Am I good hearted?", the search for evidence consistent with that proposition is reflexive; it is the search for evidence *inconsistent* with the proposition that is effortful and uncertain (Crocker, 1982; Skov & Sherman, 1986; Ward & Jenkins, 1965; Wason & Johnson-Laird, 1972).

In extensive programs of research on both anchoring and social comparison, Thomas Mussweiler has shown just how rapid and effortless such confirmatory searches can be (Mussweiler, 2003; Strack & Mussweiler, 1997). This research has demonstrated, for example, that anchoring effects are obtained in the standard anchoring paradigm through a confirmation-based heightened accessibility of anchor-consistent information. When individuals are first asked a comparative question

(e.g., "Did Mahatma Gandhi live to be more or less than 140 years old?"), they begin by asking themselves whether the target value is the correct value ("Could Gandhi have lived to 140?"). Because they tackle such questions by engaging in a confirmatory search for pertinent information, information consistent with the target value (in this case, a very aged Gandhi) is disproportionately accessible when the critical question is addressed ("How old was Gandhi when he died?") As a result, people estimate that Gandhi lived longer after first considering whether he died before or after the age of 140 than after first considering whether he died before or after the age of 9. It is the enhanced accessibility of anchor-consistent information that gives rise to the anchoring effect.

It would be hard to maintain that people spend much time or effort searching for evidence that Gandhi lived to be 140. Rather, facts consistent with an elderly Gandhi simply spring to mind. They are elicited automatically from the details of the question and the confirmatory-search routines we habitually use to examine propositions. Indeed, the magnitude of these anchoring effects are not influenced by manipulations that increase or decrease the ability to engage in effortful thought (Epley & Gilovich, 2004a; Mussweiler & Strack, 1999; Wilson, Houston, Etling, & Brekke, 1996), attesting to the rather shallow and automatic nature of these effects. The same confirmatory search process is likely to occur when we examine the possibility that we are good drivers, are kind to our coworkers, or are likely to help someone in distress. If our view of ourselves and of these behaviors are both positive, such that we engage the question by asking whether we have the capacity in question, we are likely to search for evidence that we do indeed have such capacities.

A variety of factors collude to ensure that when people search for such evidence, they are likely to find it. Because people are typically able to choose their battles, most people experience more success in life than failure. People who lack a talent for stringing words together are unlikely to write much after completing school. People with two left feet are unlikely to take many ungraceful turns on the dance floor. To be sure, people seek out challenges and engage in risky enterprises, and none of us is a stranger to failure (Lewin, Dembo, Festinger, & Sears, 1944; Csikszentmihalyi, 1990). Still, life is short, and most people choose to spend their time engaging in tasks for which they have some talent, and avoiding those at which they are inept. This has the obvious and happy result of filling our lives with more triumphs than tribulations—even if the tribulations occupy more of our attention and exert disproportionate influence on our emotional lives (Rozin & Royzman, 2001).

This also gives most people a history of success from which they can draw when searching, in confirmatory fashion, for evidence that they possess some desirable trait or ability. Many traits and abilities can be interpreted in numerous ways and research by David Dunning and his colleagues has shown that people tend to interpret them—often automatically—in self-serving ways (Dunning, Meyerowitz, & Holzberg, 1989). Someone who buys recycled paper for his home printer but tosses glass bottles and aluminum cans in the trash may be justified in

thinking that he scores high on his recycling efforts. But so would someone with the opposite tendencies. Even someone who engages in neither action but donates money to an environmental organization that advocates recycling may be similarly pleased with his efforts on behalf of recycling. More generally, if each person excels on one component of a given trait or ability and weighs that component heavily in the pertinent self-assessment, everyone can walk away feeling above average.

In other words, our most rapid mental processes are associative, and among the quickest associations to spring to mind are episodes from one's past history with the domain in question. Because we try our best to structure our lives to produce positive experiences, the associations that are most accessible are likely to be positive. And because we have some leeway in how we construe various traits and abilities, these initial positive associations will often be taken as decisive and we end up with a favorable assessment of our abilities.

But what happens when associations to past episodes of directly-relevant experience do not arise? What happens when one has little or no direct experience that speaks to the ability or trait in question? In these situations people are likely to fall back on more general beliefs about themselves and how they fare on related qualities and abilities. Ehrlinger and Dunning (2003), for example, have shown that people judge their performance on many tasks by consulting chronic self-views that strike them as relevant to the task at hand. These chronic self-views, furthermore, are likely to spring to mind as unbidden products of the associative system. So when asked to judge how we might do in a novel domain, such as running a company, we automatically reframe the question as whether or not we possess the broader traits that we believe promote success at such an enterprise, such as communication skill, a strong will, or an ability to delegate effectively.

To the extent that such chronic self-views are both accurate and germane, consulting them can lead to accurate assessments. But these broader self-views may be inappropriately applied for a number of reasons. Most important for present purposes, the general self-view that is brought to bear on the particular assessment at hand may be shaped by asking confirmatory questions, engaging in a biased search for evidence, and defining which traits are relevant in a self-serving manner—just like more specific beliefs about one's abilities.

Intention as a Proxy for Action

Perhaps the quickest assessment we make, and we make it all the time, is whether a given stimulus is positive or negative (Fazio, 2001; Zajonc, 1980). Connected to this assessment is a proto-intention, a stance toward the object as something to approach or something to exert immediate energy to avoid. This is as true of conceptual representations of, say, personal traits as it is of perceptions of physical objects like Koala bears and Gila monsters. "Stalwart," "charitable," and "lively" are traits instantly seen as positive, and yield something of an approach tendency and an impulse to link them with the self. "Weak," "stingy," and "dull," in contrast, are instantly seen as negative, and yield something of an avoidance tendency

and a desire to distance them from the self. Thus, the consideration of any trait or ability carries with it something of an intention—sometimes primitive and cursory, sometimes elaborated and strongly-held—to exhibit the trait or possess the ability when it comes to positive attributes or to disavow the trait or ability when it comes to negative attributes.

Because our intentions are arrived at so quickly and are often so unambiguous in character, they sometimes substitute as input to the real question at hand—whether one will actually exhibit the behavior in question (Kahneman & Frederick, 2002; Koehler & Poon, 2004). Because people know they would *intend* to disobey Milgram's experimenter, they believe it is likely that they *would* disobey. And because people often intend to be brave, thrifty, kind, clean, and reverent, they often believe that they possess these traits in greater-than-average abundance. It is only when the barriers to fulfilling their intentions are particularly formidable that this simple substitution is over-ridden and more deliberate analysis is engaged.

Consistent with this account, people's estimates of their likelihood of completing a number of tasks are tightly coupled to their intentions to complete them, and relatively insensitive to various situational variables that actually influence whether or not they get done (Koehler and Poon, 2004). In one experiment, for example, participants estimated the probability they would participate in a future Web-based experiment to help out a student research project. Some participants were told their participation was crucial to the student's ability to complete the project whereas others were told it would merely be helpful—a manipulation that substantially affected their intentions to participate and their estimates of how likely they were to do so. In addition, half of each group of participants were told they would receive a reminder to participate just before the experiment began, and half were not. This reminder manipulation, in marked contrast to the importance manipulation, had a minimal effect on participants' estimates of the likelihood they would participate, but a substantial effect on their actual behavior.

Kruger and Gilovich (2004) provide even more direct support for the idea that people's willingness to seize on their own good intentions—while remaining unaware or insensitive to the intentions of others—contributes to overly optimistic self assessments. They demonstrate, for example, that people exhibit more of an above average effect on traits for which one's intentions are seen as particularly relevant. Intentions strike most people as a more integral component of being "appreciative" or "kind-hearted" than "popular" or "brilliant," and people tend to assign themselves higher ratings relative to their peers on the former. This result, furthermore, holds true when other variables known to affect self-enhancement—such as trait ambiguity, desirability, and observability—are held constant. If intentions are seen as a significant determinant of whether or not one possesses a given trait, people factor in their own good intentions and conclude that they will possess the trait in greater measure than most of their peers—ignoring that others are likely to have the same intentions and thus are likely to strive just as hard to possess the trait in question.

TABLE 4.1 Participants' and Observers' Ratings of Participants' Cold-Pressor Submersion Time, Altruism, and Altruistic Intent

	Percentile estimate of		
Rater	Submersion time	Altruism	Altruistic intent
Self	41%	62%*	63%*
Observer	54%	54%	59%*

Note. Values significantly greater than average (the 50th percentile) are indicated with asterisks; from Kruger and Gilovich, 2004, p. 9.

In an experimental test of this idea, Kruger and Gilovich provided participants an opportunity to earn money for their favorite charities by the quality of their performance on a cold-pressor test. The longer they kept their hand submerged in painfully cold water, the more money they earned for their chosen charity. Afterward, participants rated their performance relative to how they thought the average participant would perform, and also rated how "altruistic" they had been in the study and how strongly they had intended to behave in an altruistic fashion—again, both relative to the average participant. As the data in Table 4.1 indicate, participants' ratings of how altruistic they had been closely matched their ratings of altruistic intent but diverged from their assessments of their actual performance. When rating the self, intentions count a great deal.

But note that this is not true when assessing others. Another group of participants watched a videotape of an individual undergoing the cold-pressor test and then made the same three ratings of that person. In this case, ratings of how altruistic the person had been were tightly coupled to the ratings of the participant's actual behavior, not the ratings of altruistic intent. When rating others, intentions count for less. Examined differently, when ratings of the target person's altruism are predicted from ratings of cold-pressor behavior and altruistic intent, regression analyses reveal a marked divergence in the views of targets and observers (see Table 4.2). The cold-pressor participants' own ratings of how altruistic they had been

TABLE 4.2 Relative Weights Placed on Submersion Time and Altruistic Intent in Judgments of Target's Altruism by Targets and Observers

	Rating	
Rater	Submersion Time	Altruistic Intent
Target	–.13	.85*
Observer	.59†	.09

Note. †$p < .10$. *$p < .05$; from Kruger and Gilovich, 2004, p. 10.

were heavily influenced by their altruistic intent and unconnected to their actual behavior. The relative weights of observers' ratings were just the opposite.

This pattern of weights can explain why we sometimes predict the behavior of others more accurately than our own. When intentions happen to be poor predictors of behavior, access to our own intentions can only serve to pollute our estimates of likely behavior. Being free of such knowledge about others allows us to base our predictions on more determinative information. In a clear demonstration of this phenomenon, Epley & Dunning (2000) asked Cornell University students how many daffodils either they or the average Cornell student would purchase during a campus fund-raising drive for the American Cancer Society. The students substantially overestimated how many they personally would buy, but their estimates of the average student were right on the money. The students certainly intended to contribute to the cause and that intention apparently got in the way of making an accurate assessment of their true likelihood of participation. These intentions, furthermore, were likely to have arisen quickly and effortlessly. A daffodil, after all, is an unambiguously positive stimulus for nearly everyone—so positive it could doubtless serve effectively alongside such stimuli as kittens and ice cream cones in affective priming experiments. And what scourge does one more quickly and effortlessly develop an intention to stamp out than cancer?

Notice that a tendency to focus on intentions when considering oneself but behavior when considering others is likely to leave people feeling even more positive about themselves when confronted with the actual selfish behavior of others. Hearing, for example, about the behavior of participants in the Milgram obedience study does not convince people that their intentions to disobey are off the mark, but rather serves as evidence that others are less compassionate and resolute than one thought. Or hearing about yet another instance of bystander nonintervention does not clue people into the difficulties of intervening in an emergency, but further strengthens preexisting beliefs about others' callousness.

This idea receives support from a follow-up experiment by Epley and Dunning (2000), one in which participants were asked to make hypothetical predictions about how much money they and the average person would donate to a charity after receiving $5 to complete an experiment. Not surprisingly, participants thought they would donate considerably more money than the average person. The participants were then shown, one at a time, the amounts donated by participants in a previous experiment who actually had been asked to donate to charity. These actual donations were considerably less generous than the hypothetical predictions, which resulted in participants making progressively lower predictions about the amount others would donate as they learned more and more about others' actual donations. Participants did not, however, alter their predictions about their own likely donations in any way, such that the difference between self and others grew as more information about actual donation levels was obtained. The reason for this was made clear from the results of a final questionnaire: participants reported that they had predicted their own behavior by taking stock of their personality traits and "character," but not the actual behavior of previous participants.

WHAT HAPPENS UPON FURTHER REVIEW?

We have discussed how a number of automatic assessments produced by System 1 lead most people down the path toward an illusion of personal strength. But fortunately for the health of our species, System 1 does not work alone. These initial assessments can be altered or overridden by the more deliberate, reflective processes of System 2. How are the shallow assessments discussed above likely to be altered by the deeper output of System 2?

Because these quick, shallow thoughts slant people's self-assessments in an overly positive direction, more reflective assessments might reign in this optimism and moderate people's judgments. It is a simple matter, certainly, to demonstrate that there is "room" for such moderation. Toward this end, Williams and Gilovich (2004a) asked one group of participants to make the standard ratings that typically elicit the above average effect—to rate themselves relative to their peers on such traits as creativity, thoughtfulness, sophistication, and so on. A second group was asked to provide, not a single percentile score, but a pair of percentile scores representing the upper and lower bounds of where they could *possibly* fall among their peers. Across these domains, the first group's mean "point estimate" fell at the 65th percentile, nearly identical to the second group's mean upper bound at the 67th percentile but considerably higher than the mean lower bound (49th percentile). The typical self-assessment is clearly one that gives the person making it the benefit of a rather substantial doubt. Where people conclude they fall on a given trait distribution is at the very highest level they believe they could possibly fall.

That the typical self-assessment is right at the subjective maximum implies that closer scrutiny might motivate people to moderate their assessments. The key questions, then, are how often do people engage in such scrutiny and how substantially does it moderate their assessments? As we shall see, there are reasons to believe that this sort of closer scrutiny is rarely engaged and, even when it is, it may only rarely lead to substantially less optimistic self-assessments.

Further Review: The Exception, Not the Rule

Fortunately for fans of the National Football League, the tedious "further review" devoted to controversial plays is only rarely enacted. And fortunately for the self-esteem of the average person, there is reason to believe that further review of one's initial, shallow assessments of personal strength may also be rarely engaged. The effortful processes of System 2 are most likely to be devoted to such assessments when there is some problem to devote them to. But when it comes to everyday self-assessments, there is typically no problem to signal the need for System 2's assistance. The reflexive confirmatory search for evidence of personal strength does not have the feel of a biased search, and the quick elicitation of one's intention likewise feels like an aide to accurate assessment, not a contributor to biased assessment. There is therefore little in these initial mental operations, or in the

product of these mental operations, to trigger corrective efforts. More generally, automatic cognition is experienced as unbiased because there is no trace of the sort of effortful mental work that people associate with biased reasoning.

The fact that many of the sources of biased self-assessment are automatic and thus inaccessible to introspection gives rise to an illusion of personal strength that folds in on itself—the illusion that we have the strength to stand up to the temptation to be biased. Given that people think they are above average on so many desired traits, it is no surprise that they tend to believe they are above average in resisting bias. Pronin, Gilovich, and Ross (2004) present data indicating that people believe they are less prone to bias than their peers in part because they tend to "look inward" to try to detect traces of bias in themselves, but consult abstract theories of bias when assessing the judgments of others. Because many biases—like those discussed here—result from processes that operate outside of awareness, the inward search for bias turns up empty. Indeed, the inward search for bias is as likely to turn up evidence that one acted or decided as one did in spite of potentially biasing considerations, not because of them (Ehrlinger, Gilovich, & Ross, 2005). People are thus frequently blind to biases in themselves that they readily detect in others.

Further Review is More Often Triggered by Negative Initial Assessments, Not Positive Ones

An intriguing program of research on motivated reasoning by Peter Ditto and his colleagues makes it clear that people are unlikely to engage in further review of their initial automatic assessments if those assessments are positive (Ditto & Boardman, 1995; Ditto & Lopez, 1992; Ditto, Munro, Apanovitch, Scepansky, & Lockhart, 2003; Ditto, Scepansky, Munro, Apanovich, & Lockhart, 1998). It is the (relatively rare) negative assessments that trigger deeper processing. The circumstances under which System 2 is activated and deeper analysis engaged, then, are not those that are likely to disabuse individuals of the notion that they have a considerable amount of ability or possess a positive trait in abundance.

Briefly, what Ditto and colleagues have found is that favorable or desirable information—"good news"—is processed effortlessly and mindlessly, whereas unfavorable or undesirable information—"bad news"—commands attention and more effortful processing (see also Dawson, Regan, & Gilovich, 2002; Gilovich, 1991; Schaller, 1992). In one noteworthy demonstration of this tendency, Ditto and Lopez (1992) gave undergraduates a test for a fictitious medical condition called "TAA deficiency" said to be associated with pancreatic disorders later in life. The test was straightforward: one simply puts one's saliva on a piece of yellow paper and observes whether it changes color within 20 seconds. In the *deficiency* condition, participants were told that if the paper remained the same color (yellow), they had the medical condition; in the *no deficiency* condition, participants were told that if the paper changed to a dark green, they had the condition. The paper remained yellow for all participants. Thus, those in the deficiency condition

would be motivated to see the paper change color and they should be disturbed by the evidence with which they were confronted—that the paper remained yellow. Indeed, these participants gave the test almost 30 seconds longer to work than those who received more welcome evidence. Participants in the no deficiency condition took in the good news and moved on; those in the deficiency condition subjected the bad news to further scrutiny.

Thus, if the sort of automatic assessments we have discussed here produce a favorable result, they are unlikely to be examined further and hence unlikely to be corrected. Instead, it is the negative initial assessments that are likely to elicit deeper, corrective processing. The net effect is that further review may be more likely to feed the illusion of personal strength than to dampen it.

Adjustments Made from Initial Assessments Are Typically Insufficient

Sometimes, of course, an initial optimistic assessment will prompt further processing and lead to the conclusion that a revision is in order. "Things aren't that rosy." "Maybe I'm being too generous with myself." "I'm not always that magnanimous." On the rare occasions when our initial, positive self-assessments *are* corrected, how are they modified? How does the correction process work? A substantial body of research indicates that initial assessments such as these are modified through a conscious, effortful, and serial process of adjustment. This is particularly true when the "anchor"—in this case, the initial self-assessment—is self-generated (Epley, in press; Epley & Gilovich, 2001; Gilovich, 2002). People typically know the proper direction in which to adjust and so this process nearly always results in assessments that are more accurate than they would otherwise be. But the adjustments people make nonetheless tend to be insufficient, leaving the final assessments off the mark (Epley & Gilovich, 2004a; Tversky & Kahneman, 1974). In this case, although deeper processing may moderate the illusion of personal strength, it is unlikely to eliminate the illusion altogether.

A pair of studies that were designed to ameliorate optimistic self assessments support this contention. In one, half of the participants were primed with honesty before being asked to fill out the sort of trait-rating scales that have regularly produced an above average effect (Williams and Gilovich, 2004b). The other participants were subjected to neutral primes and served as controls. The manipulation was effective: Those for whom the concept of honesty had been primed rated themselves less favorably than the control participants. But they still exhibited the above average effect. Across all of the positive traits, the average percentile rating among control participants was 66%. The corresponding percentile among the "honest" participants was 60%. The pull for honesty lowered participants' estimates, but it did not lower them enough. McKenna and Myers (1997) examined whether accountability would diminish the above average effect. Specifically, they asked participants to rate their driving ability relative to that of the "average driver" on

three global dimensions and 17 specific subskills. They found that accountable participants provided ratings of general ability and specific subskills that were lower than those provided by control participants. But here too the ratings of the targeted group were, except in a few cases, still above average.

The research by Kruger (1999) that we discussed earlier reinforces this point. He had participants rate their relative ability to perform a number of tasks, and he had them do so while they were either cognitively busy or unencumbered and could consider their strengths and weaknesses in greater depth. He found that the busy participants' ratings were more extreme than those provided by participants with ample cognitive resources: they rated themselves more above average on simple tasks and more below average on difficult tasks. His argument is that, typically, people seize right away on the difficulty of the task, assess their standing among their peers based on this assessment of ease or difficulty, and then adjust to take into account whether the task is likely to be easy or difficult for others as well. Individuals too preoccupied to adjust their initial assessments leave out this extra step and so their final judgments are more extreme. His data support this account but it is important to note that even participants who were not cognitively busy made rather extreme assessments. That is, these participants likewise rated themselves above average in their ability to perform simple tasks and below average in their ability to complete difficult ones. They adjusted, but their adjustments were insufficient.

Finally, consider one of daily life's more commonly encountered illusions of personal strength—the "planning fallacy," or the pervasive tendency for people to estimate that they will complete projects more quickly than they actually do (Buehler, Griffin, & Ross, 1994). From construction contractors to clerical workers, small tasks that seem like they should take only a few minutes consume the better part of a day, and week-long chores can balloon into month-long burdens. Frequent experience with this particular illusion of personal strength may enable people to recognize that their initial forecasts are likely to be overly optimistic, but eliminating this optimistic bias has proven to be extremely difficult. Indeed, even people's predictions of their completion times under a "worst-case scenario" still underestimate the time required, on average, to complete important projects (Buehler et al., 1994).

To see why the planning fallacy is so common and persistent, consider how a person typically goes about predicting when a project will be completed. First, at least some of the detailed components of the project must be considered, as well as one's estimated sustained desire to work on the project (likely to be overestimated), and maybe one or two obvious obstacles to successful completion. Frequent experience with the planning fallacy will likely lead a person to adjust an initial estimate of completion time to accommodate inevitable but unknown delays. This adjustment is likely to involve a step-by-step increase to one's initial estimate (one more day? three more days? seven more days?) until a plausible final estimate is reached. But because such adjustments stop as soon as one reaches the first plausible value, they tend to be insufficient, and the final forecasts too optimistic.

Evidence supporting the role of insufficient adjustment in the planning fallacy comes from an experiment in which college students were asked to predict how long it would take them to complete four academic projects that were due sometime during the following month (Epley & Gilovich, 2004b). Participants in one condition were asked to make these estimates while nodding their heads up and down (a nonverbal "yes") and the other half were asked to shake their heads from side to side (a nonverbal "no"). Previous research has found that people are more inclined to believe a proposition if they consider it while nodding their heads up and down than while shaking their heads from side to side (Wells & Petty, 1980). If participants in this context were adjusting from an overly optimistic initial assessment, these head movements should influence the likelihood of accepting an estimate encountered early in the process of adjustment. Those nodding their heads up and down while predicting their completion times should therefore predict that they would be done sooner than those shaking their heads from side to side. That is exactly what happened. Those nodding their heads up and down thought they would complete their projects, on average, 5.1 days before they were due, whereas those shaking their heads from side to side thought they would finish only 2.9 days early.

More important, when participants were contacted at the end of the month and asked when they actually completed each project, those who had nodded their heads when making their estimates showed a stronger planning fallacy than those who had shaken their heads from side to side. Head nodders completed their tasks, on average, 6.2 days *after* their predicted completion date, compared to only 1.4 days after among the head shakers. Here as elsewhere, therefore, it is not that people are always unaware that their shallow thoughts are likely to be overly optimistic; it's that the adjustments they make to rein in their reflexive optimism tend to be insufficient.

AUTOMATIC SELF-ASSESSMENTS NOT OPTIMISTICALLY BIASED

Our account of the quick, relatively mindless sources of optimistic self-assessments also specifies when such assessments are less likely to be so optimistic. Recall that it is a person's implicit positive self-regard that creates an automatic association between the self and positively-valued traits and abilities. This association, in turn, leads to the reflexive search for information that one possesses the positive attributes in question.

This implies, of course, that people who lack this implicit positive self-regard should be less prone to the kind of illusion of personal strength we have examined throughout this chapter. Depressed individuals are the most obvious example. Individuals who suffer from depression have been shown to be less likely to exhibit the sort of optimistic self-assessments characteristic of the nondepressed population (Tabachnik, Alloy, & Crocker, 1983; Taylor & Brown, 1988). To be sure,

many of the cognitions that accompany depression are of the elaborated, quite conscious sort that lead to destructive rumination (Nolen-Hoeksema, 2000) and are amenable to conscious intervention and retraining (Beck, 1970; 1991; Hollon, Haman, & Brown, 2002). But that is surely not all there is to depressive thought. A depressed individual's conscious self loathing may exist alongside negative self-assessments that arise automatically, and may be built up on top of such automatic self-assessments in the manner we have described here. The reflexive links between "me" and positively-valenced traits and "not me" and negatively-valenced traits that exist for the bulk of the population may be turned around in the depressed (Greenwald & Banaji, 1995). This would give rise to a shallow pessimism that parallels the shallow optimism we have examined throughout this chapter.

The present analysis also implies that the illusion of personal strength is likely to exhibit a different profile among members of collectivist cultures whose self-regard is structured differently. Members of collectivist cultures appear to have lower self-esteem, on average, than members of individualist cultures (Kitayama, Markus, Matsumoto, & Norasakkunkit, 1997), and so they may less readily form a unit between the self and positively valued traits and activities. This should lead, in turn, to self-assessments that are less optimistically biased, a pattern found with some regularity in the relevant literature (Heine, Lehman, Markus, & Kitayama, 1999; Markus & Kitayama, 1991). At the same time, members of collectivist cultures may be more likely to form a quick, reflexive unit between the groups to which they belong and various positively valued traits and abilities. This should lead to strong illusions, not of personal strength, but of in-group strength. Indeed, this exact pattern of group-enhancement in Asian cultures, compared to self-enhancement in European cultures, has been reported in the literature (Endo, Heine, & Lehman, 2000; Sedikides, Gaertner, & Toguchi, 2003).

CODA

People's insistence that they would not obey in the Milgram experiment, their conviction that they would intervene in the sorts of emergencies studied by Darley and Latane (1968), or their belief that, on average, they are very much above average, can all seem like the products of considered thought. A person takes in what seems to be the most pertinent evidence and arrives at a deliberate assessment. The attempt to explain the optimistic biases that plague such assessments thus boils down to identifying the biases, both cognitive and motivational, that plague our deliberate thoughts. What we have endeavored to explore here, however, is how optimistic self-assessments might arise from a set of more automatic mental operations. Many optimistic biases, we have tried to show, are the product of shallow thought, not reflective consideration. It is our hope, of course, that in trying to explicate the shallow thinking that underlies so many self assessments, we haven't been guilty of too much shallow thinking ourselves.

REFERENCES

Beck, A. T. (1970). *Depression: Causes and treatment.* Philadelphia: University of Pennsylvania Press.

Beck, A.T. (1991). Cognitive therapy: A 30-year retrospective. *American Psychologist, 46,* 368–375.

Bierbrauer, G. (1979). Why did he do it? Attribution of obedience and the phenomenon of dispositional bias. *European Journal of Social Psychology, 9,* 67–84.

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.

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology.* New York: Guilford Press.

Crocker, J. (1982). Biased questions in judgment of covariation studies. *Personality and Social Psychology Bulletin, 8,* 214–220.

Csikszentmihalyi, M. (1990) *Flow: The psychology of optimal experience.* New York: Harper and Row.

Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 8,* 377–383.

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

Ditto, P. H., & Boardman, A. F. (1995). Perceived accuracy of favorable and unfavorable psychological feedback. *Basic & Applied Social Psychology, 16,* 137–157.

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., Munro, G. D., Apanovitch, A. M., Scepansky, J. A., & Lockhart, L. K. (2003). Spontaneous skepticism: The interplay of motivation and expectation in responses to favorable and unfavorable medical diagnoses. *Personality & Social Psychology Bulletin*, *29*, 1120–1132.

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

Dunning, D., Meyerowitz, J. A., & Holzberg, A. (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.

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.

Ehrlinger, J., Gilovich, T., & Ross, L. (2005). Peering into the bias blindspot: People's assessments of bias in themselves and others. *Personality and Social Psychology Bulletin, 31,* 680–692.

Endo, Y., Heine, S. J., & Lehman, D. R. (2000). Culture and positive illusions in close relationships: How my relationships are better than yours. *Personality & Social Psychology Bulletin, 26,* 1571–1586.

Epley, N. (in press). A tale of tuned decks? Anchoring as accessibility and anchoring as adjustment. To appear in J. K. Koehler & N. Harvey (Eds.), *The Blackwell handbook of judgment and decision making.* Blackwell Publishers.

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.

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: An examination of self-generated and experimenter-provided anchors. *Psychological Science, 12,* 391–396.

Epley, N., & Gilovich, T. (2004a). Are adjustments insufficient? *Personality and Social Psychology Bulletin, 30*, 447–460.

Epley, N., & Gilovich, T. (2004b). *Insufficient adjustment from best- and worst-case scenarios.* Unpublished manuscript, Harvard University.

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

Fazio, R. H. (2001) On the automatic activation of associated evaluations: An overview. *Cognition & Emotion*, *15*, 115–141.

Ferguson, M. J., & Bargh, J. A. (2004). How social perception automatically influences behavior. *Trends in Cognitive Sciences, 8*, 33-39.

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

Gilovich, T. (2002). Anchoring in egocentric social judgment and beyond. In J. P. Forgas & K. D. Williams (Eds.), *The social self: Cognitive, interpersonal, and intergroup perspectives* (pp. 37–50). New York: Psychology Press.

Greenwald, A. G., Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109,* 3–25.

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

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review, 106,* 766–794.

Hollon, S. D., Haman, K. L., & Brown, L. L. (2002). Cognitive-behavioral treatment of depression. In I. H. Gotlib & C. L. Hammen (Eds.), *Handbook of depression* (pp. 383–403). New York: Guilford Press.

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

Kitayama, S., Markus, H. R., Matsumoto, H., & Norasakkunkit, V. (1997). Individual and collective processes in the construction of the self: Self-enhancement in the United States and self-criticism in Japan. *Journal of Personality & Social Psychology, 72*, 1245–1267.

Klar, Y., & Giladi, E. (1999). Are most people happier than their peers, or are they just happy? *Personality and Social Psychology Bulletin, 25,* 586–595.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypotheses testing. *Psychological Review, 94*, 211–228.

Koehler, D. J., & Poon, C. S. K. (2004). *Self-prediction based on strength of intention.* Manuscript under editorial review.

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., & Burrus, J. (2004). Egocentrism and focalism in unrealistic optimism (and pessimism). *Journal of Experimental Social Psychology, 40*, 332–340.

Kruger, J., & Gilovich, T. (2004). Actions, intentions, and self assessment: The road to self-enhancement is paved with good intentions. *Personality and Social Psychology Bulletin, 30*, 328–339.

Lewin, K., Dembo, T., Festinger, L., & Sears, P. S. (1944). Level of aspiration In J. McV. Hunt (Ed.), *Personality and the behavior disorders* (pp. 333–378). New York: Ronald.

MacDonald, T. K., & Ross, M. (1999). Assessing the accuracy of predictions about dating relationships: How and why do lovers' predictions differ from those made by observers? *Personality and Social Psychology Bulletin, 25,* 1417–1429.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98,* 224–253.

McKenna, F. P., & Myers, L. B. (1997). Illusory self-assessments — Can they be reduced? *British Journal of Psychology, 88,* 39–51.

Moore, D. A., & Kim, T. G. (2003). Myopic social prediction and the solo comparison effect. *Journal of Personality and Social Psychology, 85,* 1121–1135.

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.

Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology, 109,* 504-–511.

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review, 111,* 781-799.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality & Social Psychology Review*, 5, 296–320.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing.* Cambridge, MA: MIT Press.

Schaller, M. (1992). In-group favoritism and statistical reasoning in social inference: Implications for formation and maintenance of group stereotypes. *Journal of Personality and Social Psychology, 63,* 61–74.

Sedikides, C., Gaertner, L., & Toguchi, Y. (2003). Pancultural self-enhancement. *Journal of Personality & Social Psychology*, *84*, 60–79.

Skov, R. B., & Sherman, S. J. (1986). Information gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology, 22,* 93–121.

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

Slowiaczek, L. M., Klayman, J., Sherman, S. J. & Skov, R. B. (1992). Information selection and use in hypothesis testing: What is a good question, and what is a good answer? *Memory and Cognition, 20*, 392–405.

Smith, E. R. (1996). What do connectionism and social psychology offer each other? *Journal of Personality and Social Psychology, 70,* 893—912.

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology, 73*, 437–446.

Tabachnik, N., Crocker, J., & Alloy, L. B. (1983). Depression, social comparison, and the false-consensus effect. *Journal of Personality and Social Psychology, 45,* 688–699.

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

Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (vol. 25, pp. 331–376). New York: Academic Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Ward, W. C., & Jenkins, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology, 19,* 231–241.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content.* Cambridge, MA: Harvard University Press.

Wells, G. L., & Petty, R. E. (1980). The effects of overt head movements on persuasion: Compatibility and incompatibility of responses. *Basic and Applied Social Psychology, 1*, 219–230.

Williams, E., & Gilovich, T. (2004a). *The benefit of doubt: Motivated reasoning and the above average effect.* Unpublished data.

Williams, E., & Gilovich, T. (2004b). *Honestly, maybe I'm not so good: The effect of honesty primes on the above average effect.* Unpublished data.

Wilson, T. D., Houston, C., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General, 4*, 387–402.

Windschitl, P. D., Kruger, J., & Simms, E. N. (2003). The influence of egocentrism and focalism on people's optimism in competitions: What affects us equally affects me more. *Journal of Personality and Social Psychology, 85,* 389–408.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35,* 151–175.