

## Cognitive Optimism: Self-Deception or Memory-Based Processing Heuristics?

Janet Metcalfe

Department of Psychology  
Columbia University

*In a wide variety of cognitive tasks, people's expectations of what their own performance is, was, and will be are an overestimate of reality. After documenting a number of situations in which people exhibit such overconfidence, several reasons for it are discussed. It is suggested that in these situations, the optimistic bias is not attributable exclusively to self-deception or wishful thinking. Rather, the information yielded up by the cognitive system, in combination with the heuristics used for making judgments of future and past performance tailored to the specific questions asked of participants, produces a bias toward believing that one knows, knew, and will perform better than actual performance substantiates. Consequently, in the cognitive domain, the inflated beliefs that result in overconfidence also result in cessation of efforts before the correct solution of problems is ascertained, before accurate retrieval of memorial information is attained, or before adequate learning of new material has been accomplished. This effect seems in contrast to findings on people's moods and self-esteem in real-world or threatening situations that suggest that an optimistic bias in these domains be person protective and adaptive.*

There are many cognitive domains in which people's feelings of how well they are performing, will perform, or did perform has been studied (Metcalf & Shimamura, 1994). In most, a pattern of overconfidence is observed. People think they will be able to solve problems when they won't; they are highly confident that they are on the verge of producing the correct answer when they are, in fact, about to produce a mistake; they think they have solved problems when they haven't; they think they know the answers to information questions when they don't; they think they have the answer on the tip of their tongue when there is no answer; they think they produced the correct answer when they didn't, and furthermore, they say they knew it all along; they believe they have mastered learning material when they haven't; they think they have understood, even though demonstrably they are still in the dark.

This overconfidence effect is superimposed on what, in many domains, is a fairly accurate perception of the ordering of difficulty of problems or memory questions (Blake, 1973; Brown & Lawton, 1977; Cultice, Somerville, & Wellman, 1983; Freedman & Landauer,

1966; Gruneberg & Monks, 1974; Hart, 1965; Lachman, Lachman, & Thronesbury, 1979; Metcalf, 1986a, 1986b; Metcalf & Wiebe, 1987; Nelson, Gerler, & Narens, 1984; Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982; Schacter, 1983; Wellman, 1977). Thus, if one asks people how likely it is that they will be able to correctly recognize the answer to each of a series of questions to which they cannot recall the answers immediately, they will, in general, give higher probabilities to questions they are more likely to be able to answer and lower probabilities to questions they are less likely to be able to answer. Schwartz and Metcalf (1994) called the correspondence between rank ordering within a set and performance on individual problems the person's *micropredictive accuracy* and contrasted it with the correspondence between overall expectations and proportion correct, which they called the person's *macropredictions*. Whereas the former show that people have some ability to predict what they will know, the latter show a pattern of overconfidence.

In the first section this article, I describe some of these metacognitive illusions. Then I discuss two explanations for them: the self-deception hypothesis and the misattribution hypothesis. Although there are some domains of optimism in which self-deception may play a role, there is growing consensus that the overconfidence seen in metacognitive judgments about memory and cognitive performance stems primarily from an attribution process in which incorrect information is treated as

---

This research was supported by National Institute of Mental Health Grant MH48066.

I thank George Wolford, Walter Mischel, and Rani Bose.

Requests for reprints should be sent to Janet Metcalf, Department of Psychology, Columbia University, New York, NY 10027. E-mail: Metcalf@psych.columbia.edu.

if it were correct or is taken as a predictor of future correct performance. This research in metacognition, relying heavily on inferential processes about fuzzy information and feelings, is thus linked to the attribution-theoretic position delineated in the seminal works of Kelley (1971) and Schachter and Singer (1962) as well as to the judgmental heuristics given by Tversky and Kahneman (1974).

### Metacognitive Phenomena

#### People Think They Will Be Able to Solve Problems When They Won't

The issue of whether people know when they will be able to solve certain problems is a question of great practical as well as theoretical interest. Given that people take on problems they think they can solve, if this belief is accurate and appropriate, then effort allocation can be calibrated, and people will not attempt the impossible or the impossible for them. This kind of metaknowledge has been studied intensively with general information questions (which are addressed subsequently), but, although the issue may be more important in problem solving, it has been less frequently examined. However, there are some studies of metacognition in problem solving (e.g., Simon, 1979).

In a study that investigated the feeling of knowing in a problem-solving context, participants were given a series of insight or noninsight problems to solve (Metcalf, 1986a; Metcalf & Wiebe, 1987). An example of an insight problem is the following: A landscape gardener is given instructions to plant four special trees so that each one is exactly the same distance from each of the others. How is he able to do it? An example of a noninsight problem is this: Factor:  $16y^2 - 40yz + 25z^2$ . Those problems to which the participants knew the answers immediately were pruned from the set on the grounds that, on this subset, they might be remembering the answers rather than actually solving the problems. For the remainder of the problems, participants were asked to provide a numerical estimate of the probability that they would be able to solve that problem when given 5 min to do so. On both sets of problems, people apparently thought they would do much better than they did—their ratings exceeded their actual performance. The overconfidence effect was more pronounced on the more difficult problems, which, in this case, were the insight problems. The pattern observed in this study, of being most overconfident on the materials on which the confidence is least justified—the extremely difficult material—is a pattern that recurs in the literature on confidence judgments (Lichtenstein, Fischhoff, & Phillips, 1982; Maki & Serra, 1992).

#### People Are Highly Confident That They Are on the Verge of Producing the Correct Answer When They Are, in Fact, About to Produce a Mistake

A second study looked at people's judgments about progress on a problem-solving task while they were in the throes of doing the task (Metcalf, 1986b). Feeling-of-"warmth" judgments, indicating how close the participant thought he or she was to problem solution, were assessed every 10 sec during the course of problem solving—giving a marker of the dynamic relation between participants' moment-to-moment expectations and their actual performance. Simon and Newell's General Problem Solver (see Newell & Simon, 1972; Simon, 1979) implicated such metacognitive feelings of warmth (a term they introduced into the scientific literature) as the main guiding force that allows for non-random search in human problem solving. They suggested that to make progress in problem solving, the person must traverse a path from their current state of knowledge to a final state that is the solution to the problem. If the person were to make merely random path selections, the problem would be virtually unsolvable: It would be analogous to a chimpanzee at a keyboard accidentally, by random probing, composing Beethoven's 5th—not impossible but not likely. Some monitoring process is needed to allow a directed rather than a random search. Simon and Newell suggested that people are able to assess their distance from the solution and then use a heuristic of choosing the path in the tree structure that minimizes the distance, at every step—diverging from this strategy only when an impasse is reached. To enact this seemingly necessary problem-solving heuristic, people must be able to assess feelings of warmth relatively accurately.

In the experiments, participants were given a series of problems, including insight problems, routine problems, and anagrams and were asked to indicate how close they were to solving the problem while they were in the course of doing so. These ratings were taken as unobtrusively as possible—having participants simply make a mark on a scale every 10 sec to the sound of a beep. More verbal responding can severely interfere with the problem-solving process itself. So, for example, people were given problems like the following: "Suppose you are a museum curator, and a man comes in and offers you an ancient-looking bronze coin that is stamped 544 B.C. You have accepted objects from suspicious looking people before but in this case you immediately call the police. Why?"

The finding of most interest was that the warmth ratings were low immediately before people were about to give the correct answer but were high before people were about to produce an error of commission. The feeling that they were about to get the answer predicted errors, not answers.

One explanation for this counterintuitive finding was that perhaps people were using a *satisficing strategy* (Simon, 1979)—accepting an answer that is “good enough.” Such a strategy may be of particular value in situations in which there is no correct or perfect solution or when some decision, regardless of its quality, has to be made because of time constraints. In most situations, there are pros and cons to any number of possible solutions, and there is no absolutely correct solution. Under such conditions, it is reasonable for a person to satisfice. However, if one is solving an insight problem, where there really is a correct answer, then a satisficing strategy is not one that will typically produce a good result. Saying to oneself that bronze was not yet invented in 544 B.C. is not a particularly convincing solution, and people do not produce an aha response when they generate this response. What they apparently do is produce such a response and give it a fairly high warmth rating—being better than having no hypothesis at all, some metacognitive warmth is attached to it. Metcalfe (1986b) suggested that perhaps people sometimes convince themselves that “pretty good” is “good enough,” and the warmth ratings creep up to the threshold.

To investigate this hypothesis, participants were given the suggestion that the anagram solutions were often rare words but that past research had shown that if people guessed, they would be likely to be correct. Therefore, they should guess. Performance, both on the feeling-of-warmth ratings and on the responses made on the anagram task, was compared in a condition where participants were encouraged in a satisficing strategy with a condition in which no instructions were given. The satisficing-inducing instructions produced not only many more commission errors but also a significantly higher feeling of warmth prior to response production. The question of whether or not this thought process should be considered self-deception or whether, perhaps, a more innocent label is appropriate, is addressed later.

### **People Think They Know the Answers to Questions When They Don't**

When people are asked to say how likely it is that they will get the answer to a question correct, they show an extremely well-documented tendency to overestimate this probability. Typical calibration curves have been given by Lichtenstein, Fischhoff, and Phillips (1982). They show a consistent pattern of overconfidence: When people expect to attain a score of .8, their actual score on the test is somewhere just over .6; when they expect to attain .9, they get only about .7. Typically, the discrepancy between what score the performance should be given and what the participant's estimates are is greater when the estimates are high than

when they are more moderate, that is, they are more overconfident at the high extreme of the rating scale. On some calibration curves, a small underconfidence effect is shown with the extremely easy items. But the overwhelming bulk of the data point to a phenomenon of overconfidence.

Fischhoff, Slovic, and Lichtenstein (1985) were especially interested in the effect at the extremes. In one study, when the participants replied that they were 100% sure of getting the answer (to a question like “What is the probability that absinthe is a precious stone?”), the correct responses ranged from 71.7% to 83.1% correct (depending on the format of the criterion question). Absolute certainty does not appear to mean absolute certainty. In a related study, participants were asked to choose which of two causes of death was more frequent and to indicate confidence by giving an estimate of the odds that they were likely to be right. People were fairly well calibrated when the odds ratios given were low (up to about 3:1), but as they became high, no differences in their probability of being correct emerged. Thus, there was no difference in accuracy as a function of whether people gave their odds of being correct as 3:1 versus 100:1. The overconfidence was astonishing (perhaps indicating little more than that people did not really understand the odds scale). A total of 51% of the responses given were with odds greater than 50:1, which means that people would have to be at least 98% correct to be well calibrated. At 50:1 odds, though, people were actually correct only about 68% of the time. About one fourth of the responses were given with odds over 1,000:1.

### **People Think the Answer Is on the Tip of Their Tongue When There Is No Answer, or the ‘Answer’ Is Wrong**

An amusing illusion of a metamemory illusion is the finding that people sometimes report that an answer is on the tips of their tongues, when they not only never knew the information that is supposedly lurking just behind the veil of consciousness, but when that information does not, in fact, exist. For example, Schwartz (in press) provided participants with questions like the following: “What is the last name of the only woman to sign the Declaration of Independence?” “What is the name of the heroic innkeeper in the movie *Seems Like Old Times Again*?” “What is the name of the only living reptile that flies?” “What is the last name of the Canadian author who wrote the novel *The Last Bucket*?” Schwartz found that a remarkable number of participants would claim to be in a tip-of-the-tongue state to such questions. He argued that these feelings violate the doctrine of concordance, whereby the participant's phenomenological experience is closely related to what is retrieved, because nothing should be retrieved. Of

course, people might base the tip-of-the-tongue judgment on either the familiarity of the cue or on false retrievals that are thought to be true (or even that are known to be false but that nevertheless provide a feeling of familiarity). Jones's (1989) studies on the role of interlopers in the tip-of-the-tongue phenomenon bolster the idea that incorrect information can have a distinct impact on this metacognition.

Jones (1989) showed, consistent with Woodworth's (1929) view, that interlopers can provoke a tip-of-the-tongue state. Woodworth argued that these near misses blocked access to the correct answers but thereby put a person into a state of high familiarity combined with failure of access to the answer. In contrast, Brown and McNeill (1966) suggested that interlopers might facilitate rather than disrupt retrieval. Jones's (1989) data favored the Woodworth hypothesis, showing that the presence of near misses—for example, the presence of *secant* when the participant was trying to retrieve the word *sextant*—increased the probability of participants reporting being in a tip-of-the-tongue state.

#### **People Think, Even When Given Contradictory Feedback, that They Produced the Correct Answer and that They Knew it All Along**

Fischhoff (1975, 1977) demonstrated that if people first make a prediction about the outcome of an event and then, at some later time, are given feedback about the actual outcome and are asked what they said originally, they bias their memory for their original judgments in the direction of the actual outcome; that is, they exhibit a phenomenon called *hindsight bias* or the “knew it all along” effect. For example, Arkes, Wortmann, Saville, and Harkness (1981) provided physicians with a case history followed by laboratory investigations. The physicians were then asked to provide probabilities for four possible diagnoses. The averages given by physicians were 44%, 29%, 16%, and 11% for each of the four. However, a separate group, who were told the diagnosis first (manipulated to equally represent the four possibilities), gave probability estimates of 39%, 35%, 38%, and 31%. Notably, on the two least likely diagnoses, in particular, the physicians were unable to look at the evidence in a manner unbiased by what they “knew” to be true.

Poulton (1994) noted the difficulty that this hindsight-bias phenomenon poses for researchers who have made truly exciting and unexpected discoveries. Once the reader has read of the nature of the discovery, it seems obvious, and the magnitude and brilliance of his or her accomplishment is nearly invariably diminished by hindsight. Hoch and Loewenstein (1989) noted that whereas participants are sometimes able to discriminate easy from difficult tasks, and under some circumstances

to give a judgment such as “I would never have known that,” in the case of insight problems the answers produce such a strong hindsight effect that they overwhelm feedback information. The hindsight bias may provide a severe pedagogical hurdle in such cases, engendering in teachers a dramatic lack of empathy for the student still struggling for insight.

#### **People Believe They Have Mastered Learning Material When They Haven't**

A number of researchers have discovered systematic biases in people's judgments of learning. For example, Dunlosky and Nelson (1992) varied the time between initial study of a cue–target pair and participants' judgments of learning (i.e., an assessment of whether or not they would be able to retrieve the target 10 min later). The delayed judgments were more accurate at predicting eventual performance than were the immediate judgments, which overestimated performance. In another condition, participants were provided with the target as well as the cue before making the judgment. In this case, the judgments were inaccurate overestimates regardless of the interval. Many other examples have been documented in which the participant's judgment of learning is inflated by the immediate but transient ease of access to the target item at the time at which the judgment is requested. Participants seem to use retrieval fluency (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Bjork, 1994; Jacoby, Bjork, & Kelly, 1994) in making this judgment and to discount or ignore factors that might make that indicator a poor barometer of future performance.

#### **People Think They Have Understood, Although They Are Demonstrably Still in the Dark**

Many studies have shown that people believe they have understood text passages when their later test performance reveals that they have not or even when there are internal inconsistencies within the text that make it self-contradictory. For example, Glenberg, Wilkinson, and Epstein (1982) and Epstein, Glenberg, and Bradley (1984) had participants read passages containing internal contradictions. Even when explicitly told to find the contradictions, participants not only missed them but also voiced high confidence in their clear understanding of these inherently incomprehensible passages—exhibiting an illusion of comprehension. Similarly, when people are asked questions such as “How many animals of each kind did Moses take on the ark?” they quickly and confidently reply *two* (Bredart & Modolo, 1988; Reder & Kusbit, 1991) even though,

if pressed, those same people will unfailingly acknowledge that it was Noah and not Moses who built and sailed the ark. As in the insight problems, the incorrect but related and familiar lure slips in, seemingly unnoticed, as a good-enough place filler.

In more straightforward passages, where the task is simple reading comprehension followed by a test of that comprehension, there are also many experimental examples showing that people are overconfident in their judgments of comprehension (e.g., Maki & Berry, 1984; Shaughnessy, 1979; Waern & Askwall, 1981). In a number of studies, Glenberg and his group (e.g., Glenberg & Epstein, 1985, 1987; Glenberg, Sanocki, Epstein, & Morris, 1987) showed that calibration of comprehension is at or close to zero—people simply have no idea of what they have understood, especially when domain familiarity is factored out. Maki and Berry (1984) showed that whereas participants below the median on a performance test were not calibrated, those who were above the median showed a small degree of calibration. Maki & Serra (1992) also showed some calibration between comprehension and test performance. There appears to be agreement that participants use their self-knowledge of topic familiarity as a basis for their judgments—if they are familiar with a topic they will give high comprehension ratings that, depending on the nature of the test, may allow them some predictive accuracy (although whether this stems from actual comprehension of the text itself or is due to the participants' better inferential ability in their domains of expertise is not a settled issue).

## Explanations

### Self-Deception?

One possible explanation of the overconfidence shown by people in these situations is that they know that their answers are incorrect, but they convince themselves otherwise. Such an explanation would entail that the participants are metacognitively fundamentally accurate but engage in wishful thinking about their cognitive performance or that they are self-deceptive. The question, in part, is whether participants actually have correct and veridical metacognitions, which they then tilt toward optimism, or whether the bias shown so frequently is a direct result of the cognitive processes that underlie the judgments.

Although there may be some self-persuasion involved in the problem-solving case in which people accept an incorrect answer after they have expressed high warmth ratings, the construct of self-deception, as outlined by Fingarette (1969) (and discussed by Elster, 1979; Fischer, 1985; Goleman, 1985; McLaughlin & Rorty, 1988) is usually taken to imply much more than either an error of judgment or even of accepting a

satisficing solution. Fingarette (1969) specified three criteria given by Penelhum for self-deception: (a) belief in the face of strong evidence, (b) the participant's knowledge of the evidence, and (c) the participant's recognition of the import of the evidence. Even by Penelhum's criterion, most of the overconfidence data would not be considered self-deceptive because the participant might have no contradictory evidence at his or her disposal. Fingarette added a fourth crucial component to the paradox of self-deception:

I think it is also required that he deny, with a certain sincerity, not only the belief, but even that he sees the way the evidence points (though by the hypothesis he does see this). For, even were he to confess seeing where the evidence points and that it is decisive, and were he nevertheless to deny with a certain sincerity that he believes the conclusion to which it points, we would not necessarily consider him in self-deception. Such a case might merely present a puzzle to us. Or it could be the kind of case where [upon hearing of a disaster] a woman cannot absorb the shocking news—"She knows it's true, yet she can't believe it." What would lead us to the ... characterization [of self-deception] would be the growing belief that the truth is beginning to "seep in," that having begun to assimilate the truth, she is now fighting, with a certain success, to deny the import of the evidence or its decisiveness; in short, she is trying to keep herself persuaded that what she knows in her heart to be true is not really true. ... The crucial element which is necessary ... is the element of purposefulness. If our subject *persuades* himself to believe contrary to the evidence *in order to evade*, somehow, the unpleasant truth to which he has already seen that the evidence points, then and only then is he clearly a self-deceiver. (pp. 27–28)

To give an example of self-deception that meets the criteria outlined by Fingarette (1969), suppose a researcher believes that his data cannot be explained by any extant theory. Although this belief may be wrong, it is not, in and of itself, self-deception. But then suppose that some theorist shows the researcher, in an unambivalent way, that a particular theory can indeed account for his data. Suppose that under these conditions, the researcher tries to suppress publication of the contradictory findings, arguments, and proofs and continues to proclaim that no theory (including, explicitly, the one for which the belief-contradictory evidence was provided) can account for his data. Assuming that the researcher has managed to put himself to sleep, as Sartre (also see Fingarette, 1969; Goleman, 1985) would have it, concerning the evidence that contradicts his belief, this would be a case of self-deception. The crucial point here is that self-deception involves some kind of intentionality in suppressing the conflicting evidence that points to the truth or in keeping oneself "asleep" on this point.

Most writers on self-deception (or what Sartre called “bad faith”—*mauvais foi*) also point to a motive for this stance. “We usually assume the person is moved to self-deception as a way of attempting to evade distress, or as a way of attempting to maximize satisfaction” (Fingarette, 1969, p. 28; see also Elster, 1979; Fischer, 1985; McLaughlin & Rorty, 1988; Schelling, 1988). But motivated suppression of the truth is not at issue in simple faulty belief.

The overconfidence shown so pervasively in memory retrieval, problem-solving, and comprehension situations is probably akin to simple faulty belief and, as such, may be quite different from the motivated optimism people exhibit in other domains. Taylor and colleagues (Aspinwall & Taylor, 1992; Taylor & Brown, 1988, 1994; Taylor & Gollwitzer, 1995) showed that people in a number of potentially threatening real-life situations exhibit an unrealistic and motivated optimism that seems to border on self-deception, an optimism that she argued can be adaptive, because it can be mood and self-esteem protective. (Unrealistic motivated optimism may be different from self-deception in the strictest sense. The cancer patient who finds the silver lining in his condition may be convincing himself that he has to make the best of it. He may maintain that his chances of survival are better than they are. But the test of whether he is truly self-deceptive would occur if a cure came along. Would he hold fast to his optimistic survival estimate and abjure the need for the cure or suddenly become realistic?) It may be motivating to think that one’s chances at making a happy and successful life are better than the statistics show. Having been told that one has contracted some deadly disease, it may be helpful to believe—or make believe—that one’s chances for survival are better than they are. Such optimism, in which unpleasant or threatening outcomes are downplayed or denied, may be rightly thought to contain an element of self-deception, involving an intentionally evasive blindness to contradictory evidence in a motivated effort to protect the ego, increase coping, and reduce fear, guilt, or other negative emotions. The extent to which self-deceptive optimism is adaptive depends, as Taylor and Brown (1994) pointed out, on whether it motivates or inhibits the appropriate actions. If one’s optimism results in a failure to take the appropriate medications, then the self-deceit is obviously not adaptive. On the other hand, self-deception may keep the person from becoming overwhelmed by depression.

Although the metacognitive illusions described in this article indicate a kind of optimism, they are not, in general, a self-deceptive optimism. In none of the cases described here are the participants necessarily trying to suppress the truth—rather, they mistake some faulty information for the truth. For example, people in the insight problem experiments are typically eager to hear the correct solutions to the problems rather than being

resistant or avoidant. When told that the answer to the bronze coin problem is that no one who had actually lived in 544 B.C. could have used that date because the date itself is predicated on the supposed birthday of Christ hundreds of years later, participants immediately appreciate the explanation (often having an “aha” experience) and give up their incorrect conjectures. They do not persist in their errors in a self-deceptive and motivated manner. People in the overconfidence calibration studies of Fischhoff et al. (1985) were often (although, interestingly, not always) willing to bet on the basis of their odds judgments. It seems unlikely that they would have been willing to do this if they were using the overestimates in the service of some kind of ego gratification. Some of the participants in the experiment must have realized that their estimations were somewhat biased, however, because only 6 of the 19 were willing to play the gambling game with real money at near equal odds (i.e., when their stated odds were taken at face value, except that a very small advantage was given to the participant). Seven more participants agreed to bet but with more favorable odds. However, they did not compensate enough to break even on the bets. Six (wise and frugal) participants refused to play. Perhaps some of these 6 had overestimated for motivated reasons, but it seems more likely that they simply realized that their judgments were not accurate enough for them to risk not being able to pay for lunch.

Although there may be some advantage in some circumstances to expressing extreme certainty, in most cases the metacognitive illusions described here do not have any ego-enhancing or distress-reducing function. The hindsight bias, in which the person remembers his or her own performance as better than it was, may be a minor exception (although even this is debatable, because the possibility of discovery either by the self or others of any self-serving misrepresentation may be so hugely embarrassing and distressing to an individual that it would offset any potential urge to prevaricate). Even this bias, though, is probably more readily attributable to a simple memory blending error in which the two sources of information—participant-produced and experimenter-produced answers—are combined than to any intentional distortion of the truth.

### Memory-Based Processing Heuristics

Many researchers, studying various judgments, have converged on the idea that people base these judgments fairly globally on the information and feeling states that they have available at the time the judgment is made, whether those feeling states are attributable to information retrieved (correctly or incorrectly; see Koriati, 1993, 1994, 1995; Nelson, 1984; Nelson & Narens, 1990, 1994), to cue familiarity (see Metcalfe, 1993,

1996; Metcalfe, Schwartz, & Joaquim, 1993; Reder, 1987, 1988; Schwartz & Metcalfe, 1992), to domain or topic familiarity (see Connor, Balota, & Neely, 1992; Glenberg et al., 1982), or to ease of processing (see Begg, Armour, & Kerr, 1985; Begg et al., 1989; Bjork, 1994). The process assumed is similar to the attribution process outlined by Kelley (1971) in which causal attribution is inferred when the person detects a pattern of covariation. Although using the information available—whether correct or incorrect, relevant or irrelevant—as a guide to these kind of judgments is not an infallible method, in general it will produce a nice covariance, or a significant although imperfect correlation between the judgment and later performance. The method is also akin to the studies of Schachter and Singer (1962) in which people with ambiguous information—which, as will be discussed shortly, the retrieved information is almost bound to be—interpret that information in terms of the situation at hand. Rather than being attributions of anger or euphoria, as in Schachter and Singer's case, in the present situations they are attributions of learning, knowing, comprehension, or memory. This catch-as-catch-can use of whatever information is available appears to apply not only in metacognitive judgments but also in judgments of fame (Jacoby, Kelley, Brown, & Jasechko, 1989) or judgments of frequency (see Schwarz, 1998).

Given that people use all information at hand as if it were correct, the overconfidence bias seen in many domains can be explained easily because the information at hand is not always correct but may instead sometimes be incorrect or incomplete. Principles of human memory can be used to determine what information may be available under different circumstances and hence to ascertain the basis for metacognitive judgments and whether and under what conditions those judgments will be accurate. There is a sense in which such a heuristic approach to metacognitive judgments—an approach that is closely allied to the heuristics Tversky and Kahneman (1974) used to account for human judgments in other domains—is self-evident. Of course, people can use only the information that they have, and they have no privileged or magical access to exclusively correct information. By this view, to the extent that people do not have perfect memory, and the imperfections include both gaps and mistaken information, or even information that is close to correct but not exactly correct, their metamemory judgments are bound to be both inaccurate and systematically biased.

The answer to the question of why people tend to be overconfident in the situations detailed earlier, by this heuristic view of metacognition, is straightforward: People sometimes retrieve items or are influenced by the recollection of or fluency of items that are not quite right but are nearly right. Given that people have no supernatural knowledge of whether or not a retrieved

item is correct or whether or not the feelings that are produced are due to the correct sources, everything retrieved is assumed to be correct (see Koriat, 1993, 1994). Unless very specific precautions are taken (and even then there are limits on how much additional monitoring is possible given the constraint of the cognitive system), all sources of retrieved information will contribute to the judgments. The default attribution, then, will be that information retrieved from any source, and with any valence, is correct. However, because some of that information is incorrect—pointing to solutions other than the one designated by the experimenter—the metacognitive judgments will be inflated by these mistakes. The bias that mistakes thought to be correct will contribute to the judgments—which should be fairly pervasive across tasks—will be in the direction of overconfidence.<sup>1</sup>

### Evidence for the Heuristic View

The inadvertent acceptance of the nearly correct occurs in all kinds of guises and in all kinds of tasks, and its presence has been noted by many theorists. In this section, I review a few of the many studies that tested the hypothesis that nontarget information in memory—information that may be in some way related to but not itself be the desired answer—contributes to and inflates a variety of metacognitive judgments.

One of the earliest observations of the influence of incorrect information was by Krinsky and Nelson (1985). They analyzed feeling-of-knowing ratings as a function of the two error types: commission errors and omission errors. They found that the ratings given to errors of commission were much higher than were those given to errors of omission, suggesting that people believed that their misinformation—that which could be retrieved but was wrong—was correct.

Similarly, Oskamp (1965) conducted a study in which psychiatrists and psychiatric residents were either given a small amount of information about a hypothetical patient or were given a large amount of information. The interesting thing about this study,

<sup>1</sup>On some occasions, there is a bias toward underprediction, especially when the materials are easy. Presumably, this bias is not solely attributable to the generation of incorrect information that is thought to be correct. There are several potential causes for underestimation. First, when the materials are very easy, the participant may be aware that he or she is generating a number of distinct alternatives but that only one can be correct. This realization may lead him or her to the inference that there is trouble ahead in distinguishing which item is correct and may result in a correction of the judgment. Second, regression effects must always be considered. If the participants were merely guessing, as many researchers have noted, the result would be underestimation of performance on easy events and overestimation on difficult ones. Thus, to the extent that guessing or "noise" is a factor, we expect a reversal of the general rule of overconfidence when the materials are very easy.

from the present perspective, was that the information given (about childhood events, etc.) was nondiagnostic of the situation about which they were asked to make predictions. As the irrelevant information increased, so did participants' confidence in their answers—despite the fact that they continued to perform at chance. The irrelevant information led to an illusion of knowing.

Several studies have shown that people who have expertise in particular domains will be likely to give high predictions in the domain of their expertise. Glenberg et al. (1982) showed that people assign high feelings of knowing and comprehension to texts in domains in which they have expertise, regardless of the questions being asked. In the Glenberg et al. study, the texts were designed such that background expertise was orthogonal to comprehension. Thus, people based their judgments on domain knowledge or familiarity even when it was irrelevant. This is a pernicious illusion because it may render experts in a field—people with special knowledge in a certain domain—particularly closed minded to true advances in their own area.

In a study by Weingardt, Leonesio, and Loftus (1994), participants viewed a crime and then later received misleading suggestions about some details of the original event. They were then given a two-alternative forced-choice recognition test in which they were asked which alternative was in the original event. The interesting finding from the heuristic viewpoint was the confidence ratings on the incorrect responses that were present only in the misleading suggestions. These ratings were high; indeed, they were sometimes higher than the confidence ratings on the correct responses themselves.

Metcalf, et al. (1994) used a classic interference-theory paradigm to pit a manipulation that enhanced correct retrieval of the target against one that made performance-irrelevant information—the cue—more accessible. Feeling-of-knowing judgments followed the familiarity of the cue rather than the retrievability of the target. The influence of irrelevant, rather than relevant, information on metacognitions has been shown in several other studies as well. Reder (1987) and Schwartz and Metcalfe (1992) conducted studies in which key words in the cue were primed. For example, participants might have seen the words *Prime Minister* in a priming task prior to a general information test. This exposure increased feeling of knowing to the question "What was the name of Canada's first prime minister?" even though it did not influence the probability that the participant would be able to correctly choose *MacDonald*. Schwartz and Metcalfe found that this cue priming, which was irrelevant to the production of the correct answer, had a greater effect on metacognitive judgments than manipulations that increased the probability of correctly retrieving the target. Reder and Ritter (1992) found that participants thought they knew the answers to arithmetic questions such as  $45 \times 56$  when

they were primed on pseudo-related questions such as  $45 + 56$ .

Finally, merely repeating a statement not only inflates judgments of familiarity, of feeling of knowing, and of comprehension but it also makes people think it is true (Arkes et al., 1981; Hasher, Goldstein, & Toppino, 1977). For example, following Hasher et al., (1977), Gigerenzer (1984) presented a sample of normal residents of Schwabing (culled from the public telephone directory) with a set of *true* and *false* statements, either once or three times, and had them rate each assertion for its validity. The judged validity increased with repetition of the statements, regardless of whether or not they were actually true. Bacon (1979) contributed an interesting twist to the work on judgments of validity. He found that the judged validity of contradictory statements could benefit from the earlier presentation of the statements that they contradicted. This occurred if participants rated the contradictory statements as having been repeated, whereas if they did not make the mistake of thinking that they were repeated and instead noted the change, the contradictory statements were judged as relatively false. Thus, under some circumstances, presentation of the truth might actually increase belief in its contradiction. Similarly, Begg et al. (1985) showed that familiarity with the topic of a sentence leads to increases in ratings of validity of the sentence—an illusion of truth.

## Conclusion

In this article I provided a review of a number of overconfidence errors in human cognition, and I argued that this cognitive optimism is different from that seen in terminal cancer patients or even when one questions students regarding their prospects of attaining good grades or having a successful career. The optimism in the former cases appears to be a result of the use of an attribution process for making metacognitive judgments that relies on processing heuristics that are likely to produce a faulty—and overoptimistic—result rather than ego-protective and ego-motivated self-deception. Whereas the optimism seen in people's responses to their life situation is likely to be adaptive, the overconfidence exhibited in cognitive situations is liable to lead to premature cessation of problem-solving efforts, insufficient checking of memory retrieval, resulting in poorer performance than might otherwise be achieved, faulty assessment of the difficulty of problems for other people due to hindsight biases, insufficient study, and an inappropriate and self-defeating lack of perseverance under difficult cognitive conditions. With all these possible negative outcomes, one is led to wonder whether or not these biases can be overcome.

Glenberg et al. (1987) showed that comprehension calibration could be increased by giving a test that



allows self-generated feedback. Bredart and Modolo (1988) showed that the Moses illusion could be overcome by using a technique of focalization—where special prominence, in their case because of the structure of the sentence, was given to the discordant noun. Koriat, Lichtenstein, & Fischhoff (1980) showed that overconfidence in people's judgments about the correctness of their knowledge could be diminished if the participants were encouraged to "spell out all the reasons that you can find favoring and opposing each of the answers." The reasons opposing the participants' choice were especially important. Koriat and Goldsmith (1996) showed that when people are allowed to volunteer or withhold information, the monitoring process engaged when accuracy is emphasized allows enhanced accuracy. Metcalfe (1985b) found that people had lower presolution warmth ratings and better task performance when they were not encouraged to guess than when they were. Hoch (1985) found that predictive accuracy about results of a future job search increased when people were asked to generate pro and con reasons for their predictions. Although the pro reasons did not influence the judgments (suggesting that was what they were using anyway), the con-reason judgment task resulted in more realistic judgments. These and other examples that indicate that people can indeed overcome or at least modulate their overconfident bias provide grounds for optimism.

## References

- Arkes, H. R., Wortmann, R. L., Saville, P. D., & Harkness, A. R. (1981). Hindsight bias among physicians weighing the likelihood of diagnoses. *Journal of Applied Psychology*, 66, 252–254.
- Aspinwall, L. G., & Taylor, S. E. (1992). Modeling cognitive adaptation: A longitudinal investigation of the impact of individual differences and coping on college adjustment and performance. *Journal of Personality and Social Psychology*, 63, 989–1003.
- Bacon, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 241–252.
- Begg, I., Armour, V., & Kerr, T. (1985). On believing what we remember. *Canadian Journal of Behavioral Science*, 17, 199–214.
- Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, 28, 610–632.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–206). Cambridge, MA: MIT Press.
- Blake, M. (1973). Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 12, 311–319.
- Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, 67, 135–144.
- Brown, A. L., & Lawton, S. C. (1977). The feeling of knowing experience in educable retarded children. *Developmental Psychology*, 13, 364–370.
- Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 5, 325–337.
- Connor, L. T., Balota, D. A., & Neely, J. H. (1992). On the relation between feeling of knowing and lexical decision: Persistent subthreshold activation or topic familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 544–554.
- Cullice, J. C., Somerville, S. C., & Wellman, H. M. (1983). Preschoolers' memory monitoring: Feeling of knowing judgments. *Child Development*, 54, 1480–1486.
- Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, 20, 374–380.
- Elster, J. (1979). *Ulysses and the Sirens*. Cambridge, England: Cambridge University Press.
- Epstein, W., Glenberg, A. M., & Bradley, M. M. (1984). Coactivation and comprehension: Contribution of text variables to the illusion of knowing. *Memory & Cognition*, 12, 355–360.
- Fingarette, H. (1969). *Self deception*. London: Routledge & Kegan Paul.
- Fischer, W. F. (1985). Self-deception: An empirical-phenomenological inquiry into its essential meaning. In A. Giorgi (Ed.), *Phenomenology and psychological research* (pp. 118–154). Pittsburgh, PA: Duquesne University Press.
- Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288–299.
- Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 349–358.
- Fischhoff, B., Slovic, P., & Lichtenstein, S. (1985). Knowing with certainty: The appropriateness of extreme confidence. In M. R. Oakes & K. R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (pp. 397–415). Cambridge, England: Cambridge University Press.
- Freedman, J. L., & Landauer, T. K. (1966). Retrieval of long-term memory: "Tip of the tongue" phenomenon. *Psychonomic Science*, 4, 309–310.
- Gigerenzer, G. (1984). External validity of laboratory experiment: The frequency–validity relationship. *American Journal of Psychology*, 97, 185–195.
- Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 702–718.
- Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, 15, 486–514.
- Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, 116, 119–136.
- Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, 10, 597–602.
- Goleman, D. (1985). *Vital lies, simple truths*. New York: Simon & Schuster.
- Gruneberg, M. M., & Monks, J. (1974). Feeling of knowing and cued recall. *Acta Psychologica*, 38, 257–265.
- Hart, J. T. (1965). Memory and the feeling of knowing experience. *Journal of Educational Psychology*, 56, 208–216.
- Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16, 107–112.
- Hoch, S. J. (1985). Overconfidence in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 719–725.
- Hoch, S. J., & Loewenstein, G. F. (1989). Outcome feedback: Hindsight and information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 605–619.
- Jacoby, L. L., Bjork, R. A., & Kelley, C. M. (1994). Illusions of comprehension and competence. In D. Druckman & R. A. Bjork

- (Eds.), *Enhancing Human Performance, III* (pp. 57–80). Washington, DC: National Academy Press.
- Jacoby, L. L., Kelley, C. M., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influence of the past. *Journal of Personality and Social Psychology*, 56, 326–338.
- Jones, G. V. (1989). Back to Woodworth: Role of interlopers in the tip-of-the-tongue phenomenon. *Memory & Cognition*, 17, 69–76.
- Kelley, H. H. (1971). Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 1–26). Morristown, NJ: General Learning Press.
- Koriat, A. (1993). How do we know what we know? The accessibility model of feeling of knowing. *Psychological Review*, 100, 609–639.
- Koriat, A. (1994). Memory's knowledge of its own knowledge: The accessibility account of the feeling of knowing. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 115–136). Cambridge, MA: MIT Press.
- Koriat, A. (1995). Dissociating knowing from feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, 124, 311–333.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103, 490–517.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 107–118.
- Krinsky, R., & Nelson, T. O. (1985). The feeling of knowing for different types of retrieval failure. *Acta Psychologica*, 58, 141–158.
- Lachman, J. L., Lachman, R., & Thronesbury, C. (1979). Metamemory through the adult life span. *Developmental Psychology*, 15, 543–551.
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgments under uncertainty: Heuristics and biases* (pp. 306–334). New York: Cambridge University Press.
- Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 663–679.
- Maki, R. H., & Serra, M. (1992). The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 116–126.
- McLaughlin, B. P., & Rorty, A. O. (1988). *Perspectives on self-deception*. Berkeley: University of California Press.
- Metcalfe, J. (1986a). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 288–294.
- Metcalfe, J. (1986b). Premonitions of insight predict impending error. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 623–634.
- Metcalfe, J. (1993). Novelty monitoring, metacognition, and control in a composite holographic associative recall model: Implications for Korsakoff amnesia. *Psychological Review*, 100, 3–22.
- Metcalfe, J. (1996). Metacognitive processes. In E. L. Bjork & R. A. Bjork (Eds.), *The handbook of perception and cognition: Vol. 10. Memory* (pp. 383–411). San Diego, CA: Academic.
- Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 851–861.
- Metcalfe, J., & Shimamura, A. (1994). *Metacognition: Knowing about knowing*. Cambridge, MA: MIT Press.
- Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition*, 15, 238–246.
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95, 109–133.
- Nelson, T. O., Gerler, D., & Narens, L. (1984). Accuracy of feeling of knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General*, 113, 282–300.
- Nelson, T. O., Leonesio, R. J., Shimamura, A. P., Landwehr, R. S., & Narens, L. (1982). Overlearning and the feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 279–288.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation*, 26, 125–171.
- Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–26). Cambridge, MA: MIT Press.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Oskamp, S. (1965). Overconfidence in case-study judgments. *The Journal of Consulting Psychology*, 29, 261–265.
- Poulton, E. C. (1994). *Behavioral decision theory: A new approach*. Cambridge, England: Cambridge University Press.
- Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, 19, 90–138.
- Reder, L. M. (1988). Strategic control of retrieval strategies. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 227–260). San Diego, CA: Academic.
- Reder, L. M., & Kusbit, G. W. (1991). Locus of the Moses illusion: Imperfect encoding, retrieval, or match? *Journal of Memory and Language*, 30, 385–406.
- Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 435–452.
- Schachter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69, 379–399.
- Schacter, D. L. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 39–54.
- Schelling, T. C. (1988). The mind as a consuming organ. In D. E. Bell, H. Raiffa, & A. Tversky (Eds.), *Decision making: Descriptive, normative and prescriptive interactions* (pp. 343–357). Cambridge, England: Cambridge University Press.
- Schwartz, B. L. (in press). Illusory tip-of-the-tongue states. *Memory*.
- Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1074–1083.
- Schwartz, B. L., & Metcalfe, J. (1994). Methodological problems and pitfalls in the study of human metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 93–114). Cambridge, MA: MIT Press.
- Schwarz, N. (1998). Accessible content and accessibility experiences: The interplay of declarative and experiential information in judgment. *Personality and Social Psychology Review*, 2, 87–99.
- Shaughnessy, J. J. (1979). Confidence-judgment accuracy as a predictor of test performance. *Journal of Research in Personality*, 13, 505–514.
- Simon, H. A. (1979). *Models of thought*. New Haven, CT: Yale University Press.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103, 193–210.
- Taylor, S. E., & Brown, J. D. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological Bulletin*, 116, 21–27.

- Taylor, S. E., & Gollwitzer, P. M. (1995). Effects of mindset on positive illusions. *Journal of Personality and Social Psychology*, 69, 213–236.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Waern, Y., & Askwall, S. (1981). On some sources of metacomprehension. *Scandinavian Journal of Psychology*, 22, 17–25.
- Weingardt, K. R., Leonesio, R. J., & Loftus, E. L. (1994). Viewing eyewitness research from a metacognitive perspective. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 157–184). Cambridge, MA: MIT Press.
- Wellman, H. M. (1977). Tip of the tongue and feeling of knowing experiences: A developmental study of memory monitoring. *Child Development*, 48, 13–21.
- Woodworth, R. S. (1929). *Psychology* (2nd Rev. ed.). New York: Holt.

